Computer-Supported Collaborative Learning

Sadhana Puntambekar
Gijsbert Erkens
Cindy Hmelo-Silver *Editors*

Analyzing Interactions in CSCL

Methods, Approaches and Issues

Springer

Analyzing Interactions in CSCL

COMPUTER-SUPPORTED COLLABORATIVE LEARNING

The *Computer-Supported Collaborative Learning Book Series* is for people working in the CSCL field. The scope of the series extends to 'collaborative learning' in its broadest sense; the term is used for situations ranging from two individuals performing a task together, during a short period of time, to groups of 200 students following the same course and interacting via electronic mail. This variety also concerns the computational tools used in learning: elaborated graphical whiteboards support peer interaction, while more rudimentary text-based discussion forums are used for large group interaction. The series will integrate issues related to CSCL such as collaborative problem solving, collaborative learning without computers, negotiation patterns outside collaborative tasks, and many other relevant topics. It will also cover computational issues such as models, algorithms or architectures which support innovative functions relevant to CSCL systems.

The edited volumes and monographs to be published in this series offer authors who have carried out interesting research work the opportunity to integrate various pieces of their recent work into a larger framework.

For other titles published in this series, go to
www.springer.com/series/5814

Sadhana Puntambekar • Gijsbert Erkens
Cindy E. Hmelo-Silver
Editors

Analyzing Interactions in CSCL

Methods, Approaches and Issues

Springer

Editors
Sadhana Puntambekar
University of Wisconsin
Madison, WI, USA
puntambekar@education.wisc.edu

Cindy E. Hmelo-Silver
Rutgers University
New Brunswick, NJ, USA
cindy.hmelo-silver@gse.rutgers.edu

Gijsbert Erkens
Research Centre Learning in Interaction
Utrecht University
Utrecht, The Netherlands
g.erkens@fss.uu.nl

ISBN 978-1-4419-7709-0 e-ISBN 978-1-4419-7710-6
DOI 10.1007/978-1-4419-7710-6
Springer New York Dordrecht Heidelberg London

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Dedicated to our accomplished children
Aakash Puntambekar
Dr. Laura Hmelo
Dr. Samantha Hmelo

Acknowledgements

Like any edited volume, this book is a product of the combined efforts of several people. First, this book would not have been possible without the contribution of our colleagues who wrote the chapters. We learned from reading about the different perspectives the authors presented, and hope that our readers will do the same. We would especially like to thank our colleagues who wrote the commentaries for the book. Stephanie Teasley, Susanne Lajoie and Kristin Lund wrote the three commentaries, often at a short notice and when they had several other commitments. We thank them for their time and their valuable insights. Finally, we thank our students – Samaa Abdurraqib, Anushree Bopardikar and Dana Gnesdilow for their help with proofreading and formatting the chapters.

Introduction

Sadhana Puntambekar, Gijsbert Erkens, and Cindy E. Hmelo-Silver

Technology-enhanced environments to support student learning are becoming ubiquitous in both formal and informal educational contexts. Often, these environments require groups of students to learn collaboratively. As groups of learners engage in joint construction of meaning on a shared task, there is an emphasis on understanding how the group as a whole constructs knowledge through joint activity; this is a distinct shift from the traditional lens that views learning as a highly individualistic process and product. As Stahl et al. (2006) point out:

> Earlier studies of learning in groups treated learning as a fundamentally individual process. The fact that the individuals worked in groups was treated as a contextual variable that influenced the individual learning. In CSCL, by contrast, learning is also analyzed as a group process; analysis of learning at both the individual and the group unit of analysis are necessary. (p. 411)

This is what makes CSCL methodologically unique. This uniqueness is reflected in the several approaches that have been put forth to document and analyze collaborative interactions. CSCL as a field has made great strides from early research that focused on the extent of participation (De Wever et al. 2006). Currently, researchers use an array of qualitative and quantitative methods, including content analysis, social network analysis, analysis of log files, multilevel models, visual representations of data, etc., to analyze and model collaborative learning. Methods for analysis have included both an analysis of the process of learning and the learning outcomes. Further, measures of individual learning and learning by the group as a whole have been used.

This book is an attempt to discuss a representative set of current methods to analyze collaborative interactions, both at the individual and group levels. CSCL research tends to span across several disciplines such as education, psychology, computer science and artificial intelligence, bringing a diverse set of methods from research in these fields. The 15 chapters in this book present these diverse perspectives to provide researchers with a collection of methodologies to document and analyze collaborative interactions. A couple of recurring themes can be found through several of the chapters: unit of analysis, grain size of data, segmenting of data and temporality of interactions in CSCL. Additionally, several authors present frameworks that use multiple data sources and multiple methods of analysis.

One of the most important challenges of assessing collaborative learning is the issue of *unit of analysis*. Stahl et al. (2006) have pointed out that CSCL researchers are confronted with the issue of determining the appropriate unit for analysis. Establishing a unit of analysis poses difficulty because although group interactions are influenced by what the individual participants bring to the group, group processes are more than the sum of parts, and need to be understood as an entity within themselves. As Reimann (2007) described, learning in CSCL environments occurs "in individuals in the form of learning and in groups in the form of participation and knowledge building" (p. 611). Therefore, interactions can be analyzed with both the individual and the group as units of analysis. Further, within each level, the *grain size* of the unit needs to be determined based on the research questions that drive the analysis in a particular study (Chi 1997). Grain sizes can vary from analyzing a set of single utterances, chunks of discourse segmented along topics or themes, or both (e.g., Ash 2007; De Wever et al. 2006). These issues of units of analyses, grain size, and segmentation are addressed by several chapters in this volume.

Another recurring theme addressed in the book is that of *temporality* of data in CSCL environments. CSCL interactions occur over a period of time. Therefore analyzing single episodes does not adequately provide information about the process of learning. As Mercer (2008) described, "the coherence of educational experience is dependent on talk among participants, and so analyses of the ways that their continuing shared experience is represented and the ways that talk itself develops and coheres over an extended period are required" (p. 55).

The chapters are divided into three parts, as discussed in the next few paragraphs.

Part I: Understanding Group Processes

Kapur Voiklis, & Kinzer address the issue of uncovering temporal patterns in CSCL interactions by using a complex systems approach to the study of convergence in groups. In doing so, they address another significant aspect of group processes, that of divergence of ideas among group members and convergence of a group's understanding (Teasley et al. 2008; Roschelle 1992). Stahl (2004) argues that initial divergent ideas between group members significantly affect collaboration, because group members have to negotiate towards shared meaning. But it is not clear whether initially divergent groups eventually converge, and whether convergence is desirable. Nonetheless, despite the importance given to intersubjective meaning-making in the CSCL literature (Suthers 2006), convergence remains a difficult parameter to analyze. In their analysis, Kapur, Voiklis and Kinzer coded each discourse move in relation to the group goal, based on whether or not it moved the group toward a correct solution to the problem. A Markov model was then used to predict the group's performance.

Law, Yuen, Wong, and Leng discuss an approach to understand learners' trajectories in a group during asynchronous collaboration. They report a study in which participation statistics were combined with specific aspects of discourse data in Knowledge Forum®. Law et al. propose a methodology to enable automatic coding and visualization of productive discourse threads on three interrelated aspects (scaffolds used, argumentative discourse markers and content topics).

Collaborative learning happens in a context, and several contextual variables affect a group's learning. Context can be broadly conceived as the physical and psychological variables that emerge from person-to-person interactions in any interpersonal human environment. Broad cultural influences of family and state intermix with more local cultures of schools and peer groups to provide additional constraints on how CSCL will emerge as a context for each individual learner. Arvaja addresses this issue by proposing a methodology based on sociocultural theories of learning, taking into account how physical and contextual aspects of any environment affect a group's discourse.

An important aspect of context is the tool that facilitates collaboration. Each tool has different affordances and has to be adapted to the context in which it is being used. Hmelo-Silver and Liu focus on the notion of how the effect of computer tools are important in mediating a group's discourse by visually representing data and taking into consideration the chronological relationship between talk and tool use. Their method uses a chronological representation of data, addressing the issue of temporality of CSCL interactions. Finally, Stahl argues for studying group cognition in CSCL and presents a case from the Math Forum project. As an example, he shows how proposals structure the temporal flow of the group interaction and thereby establish the social order of group cognition. Besides a temporal dimension, a problem space of shared knowledge artifacts and an interaction space of positioned actors are co-constructed by collaborative small-groups which define other dimensions of this social order. These group processes are, according to Stahl, not analyzable as individual behaviors, but can only be understood taking the group as unit of analysis.

Part II: Understanding Learning Within Groups

As mentioned earlier, CSCL interactions occur both at the level of the individual and groups. Contributions of individual group members influence themselves, other group members, and group processes as a whole. The ways in which individuals take up ideas and how the group as a whole moves forward are important aspects of CSCL to document and analyze, especially because each of these levels is unique for any group and also for the same group at different times. As such, it is important to study how membership in a group affects an individual member's learning, as well as the temporal aspects of how this learning changes over time. Both of these issues are addressed in the chapters in this part of the book.

Two of the chapters in this part focus on how the effect of individual membership in groups can be analyzed using multilevel models. Jansen, Erkens, Kirschner, and Kanselaar explain the use of multilevel modeling to account for both the individual and group level variables in the analysis of CSCL interactions. They discuss three problems in analyzing data: hierarchically nested datasets, non-independence and differing units of analysis. They then illustrate strategies to address these problems through three examples. Stylianou-Georgiou, Papanastasiou, and Puntambekar further develop the idea of using multilevel modeling to analyze nested data by modeling the dependencies in their data to understand relationships between the variables of interest. In their study, they use both individual and group measures to apply a two-level model to understand the role of group membership in individual students' learning outcomes. Their analysis allows them to understand how attributes of the learning environment interact with group measures to affect individual learning outcomes.

Reimann, Yacef, and Kay address the issue of temporality by discussing how log data can be analyzed using data mining techniques. CSCL researchers often collect data in the form of log files to understand group interactions, resulting in large amounts of log data that need to be reduced, organized and analyzed. Log data often capture interactions that occur over time, such that events and sequences are related to each other. Reimann et al. address this issue by proposing data mining techniques that aim to identify sequence patterns and discrete event models. They propose analyzing group processes at an atomistic level as well as a holistic level. The chapter also acknowledges the challenges in applying data mining techniques, particularly the quality of data that is collected, and the level of granularity of that data, because "the quality of a model depends on the quality of the data" (Reimann et al. this volume). Thus, although data mining can be a powerful mechanism to analyze large quantities of log data, it is important to keep in mind the complexity of collecting the data in the first place and interpreting the resulting models.

The theme of temporal analysis continues in the chapter by Jeong, Clark, Sampson, and Menekse, who also propose the use of sequential analysis for group data. However, they use this approach with a coding of the discourse moves. Using a coding scheme for argumentative discourse moves, the sequential analysis helped them to identify, visualize, and assess the dialogic, temporal processes of argumentation in online science learning environments.

Part III: Frameworks for Analyzing Interaction in CSCL

The final part of this volume focuses on frameworks for analyzing collaborative interactions. Stegmann and Fischer present a model with heuristics for segmentation and coding. They discuss the challenging issue of segmenting, which is a key component of analyzing collaborative interactions. The grain size of the unit of analysis is an issue that all CSCL researchers grapple with, as we saw in earlier parts. Smaller segments in data provide finer grained analysis but little contextual information. On the other hand, larger units of analysis help create context but with the loss of detail. Therefore data segments need to be determined based on the research questions and goals for analysis (Chi 1997; Chavajay & Rogoff 2002).

Understanding group functioning as a whole, is a focus for several authors in this part of the book. This is a pertinent issue, because the quality of a group's interactions, often affect the outcomes. Groups that are dysfunctional may not accrue the benefits of learning collaboratively (e.g., Barron 2003). Rummel, Meier, Spada, Kahrimanis, and Avouris discuss the value of analyzing collaborative interactions as a whole, based on a few parameters that rate the quality of how well groups collaborate, such as communication, collaboration, etc. This is significant because failure of productive collaboration among group members can have a detrimental effect on both individual learning and collective knowledge development. Rummel et al. provide us with a tool that can be adapted and used by CSCL researchers, and can be flexibly combined with other measures, such as coding the utterances at a fine-grained level.

In a similar vein, Gweon, Jun, Lee, Finger, and Rosè tackle the difficult issue of the level and quality of interactions in face-to-face communication in groups. Difficulties in the interactional processes of groups often affect the outcomes of group work. The authors present their approach to tracking group progress in an instructional context. In a series of studies, they first identified process categories that needed to be tracked and then coded these during group work. The authors offer ways to automate this process using machine learning techniques. The approach that Gweon et al. discuss has implications for instructors and teachers to gain insight into group functions in classroom contexts. Suthers and Medina address a major topic of recent CSCL research, that of combining multiple logfiles of collaborative activities from different media and tools for analysis. They introduce the notion of contingency graphs that allow researchers to combine data that is distributed across media, enabling them to have a single abstract artifact with links to the original data.

Extending frameworks to the use of mobile devices, Scanlon describes case studies to showcase her approach to analyzing collaborative learning in several projects. Her framework, CIAO (Context, Interaction, Attitude and Outcomes) uses data collected from a variety of sources, both qualitative and quantitative. It is interesting to note that Scanlon discusses challenges to this approach as collaborative learning extends to mobile devices, creating a richer, social and technological setting. An Activity Theory framework seems to be promising to analyze the way activities are mediated by technology, as was illustrated in one of the presented studies. However, for the analysis of temporal aspects of knowledge construction a broader socio-cultural analytic approach is suggested by the author. Finally, Martínez-Monés, Harrer, Dimitriadis report on the requirements for computers to support CSCL researchers in conducting interaction analysis. The chapter notes limitations in trying to conduct analysis post-hoc in existing tools. The authors propose a design process for the development of CSCL environments. For example, a *co-design* approach in which learning and analysis needs are integrated from the start or a *multi-perspective* approach in which these two needs are treated independently at an initial stage and integrated later can be employed. Technology requirements and solutions to support this integration of learning and analysis are discussed as well.

Conclusion

Together, the chapters in this book provide a suite of tools that can be applied, modified and customized to document and analyze collaborative interactions. There are of course issues that still need to be explored. For example, while we have a range of methods for assessing learning outcomes and group processes, the issue of measuring group outcomes as a whole still remains a challenge (see Lund this volume). Our hope is to start the conversation on the different methods discussed, as the CSCL community moves forward to find the best ways to understand individual learning and group processes in collaborative environments.

References

Ash, D. (2007). Using video data to capture discontinuous science meaning making in non-school settings. In R. Goldman, R. Pea, B. Barron, & S. Derry (Eds.), *Video research in the learning sciences*. Mahwah: Erlbaum.

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

Chavajay, P., & Rogoff, B. (2002). Schooling and traditional collaborative social organization of problem solving by Mayan mothers and children. *Developmental Psychology*, *38*(1), 55–66.

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

De Wever, B., Schellens T., Valcke, M., & Van Keer H. (2006). Content analysis schemes to analyze transcripts of online asynchronous discussion groups: A review. *Computers & Education*, *46*(1), 6–28.

Mercer, N. (2008). The seeds of time: why classroom dialogue needs a temporal analysis. *Journal of the Learning Sciences*, *17*(1), 33–59.

Mercer, N., Littleton, K., & Wegerif, R. (2004) Methods for studying the processes of interaction and collaborative activity in computer-based educational activities. *Technology, Pedagogy and Education*, *13*(2), 193–209.

Reimann, P. (2007). Time is precious: Why process analysis is essential for CSCL (and also can help to bridge between experimental and descriptive methods). In C. Chinn, G. Erkens, & S. Puntambekar (Eds.), *Minds, minds, and society. Proceedings of the Computer-supported Collaborative Learning Conference (CSCL 2007)* (pp. 598–607). New Brunswick: International Society of the Learning Sciences.

Roschelle, J. (1992). Learning by collaborating: Convergent conceptual change. *The Journal of the Learning Sciences*, *2*(3), 235–276.

Stahl, G. (2004) Building collaborative knowing- elements of a social theory of learning. In J. W. Strijbos, P. A. Kirschner, & R. L. Martens (Eds.), *What we know about CSCL in higher education* (pp. 53–86) Dordrecht: Kluwer.

Stahl, G., Koschmann, T., & Suthers, D. D. (2006). Computer-supported collaborative learning: An historical perspective. In R. K. Sawyer (Ed.), *Cambridge handbook of the learning sciences* (pp. 409–426). Cambridge: Cambridge University Press.

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

Teasley, S. D., Fischer, F., Weinberger, A., Stegmann, K., Dillenbourg, P., Kapur, M., et al. (2008). Cognitive convergence in collaborative learning. In *Proceedings of the 8th international conference on the learning sciences* (Vol. 3, pp. 360–367). Mahwah: Lawrence Erlbaum.

Contents

Contributors

Maarit Arvaja
Institute for Educational Research, University of Jyväskylä, P.O. Box 35, Jyväskylä 40014, Finland

Nikolaos Avouris
Human-Computer Interaction Group, E&CE Department, University of Patras, Greece

Ellina Chernobilsky
Caldwell College, Caldwell, NJ, USA

Douglas B. Clark
Department of Teaching and Learning, Vanderbilt University, Nashville, TN, USA

Anne Deiglmayr
Institute of Psychology, University of Freiburg, Freiburg, Germany

Yannis Dimitriadis
GSIC-EMIC Group, University of Valladolid, Valladolid, Spain

Gijsbert Erkens
Research Centre Learning in Interaction, Utrecht University, Utrecht, The Netherlands

Susan Finger
Human–Computer Interaction Institute, Carnegie Mellon University, Pittsburg, PA 15213-3891, USA

Frank Fischer
Ludwig Maximilian University of Munich, München, Germany

Gahgene Gweon
Human–Computer Interaction Institute, Carnegie Mellon University, 5000 Forbes Ave., Pittsburg, PA 15213-3891, USA

Andreas Harrer
Catholic University of Eichstätt, Germany

Cindy E. Hmelo-Silver
Graduate School of Education, Rutgers University, 10 Seminary Place, New Brunswick, NJ, USA

Jeroen Janssen
Research Centre Learning in Interaction, Utrecht University, Utrecht, The Netherlands

Allan Jeong
Associate Professor, Florida State University, College of Education, Department of Educational Psychology & Learning Systems, Instructional Systems Program

Rebecca Jordan
Rutgers University, New Brunswick, NJ, USA

Soojin Jun
Human–Computer Interaction Institute, Carnegie Mellon University, Pittsburg, PA, USA

George Kahrimanis
Human-Computer Interaction Group, E&CE Department, University of Patras, Patras, Greece

Gellof Kanselaar
Centre for Learning Sciences and Technologies, Open University, The Netherlands

Manu Kapur
National Institute of Education, 1 Nanyang Walk, Singapore 637616

Judy Kay
University of Sydney, Australia

Charles K. Kinzer
Columbia University, New York, NY, USA

Paul A. Kirschner
Centre for Learning Sciences and Technologies, Open University, The Netherlands

Susanne P. Lajoie
Department of Educational and Counselling Psychology, McGill University

Nancy Law
University of Hong Kong

Joonhwan Lee
Human–Computer Interaction Institute, Carnegie Mellon University, Pittsburg, PA, USA

Jing Leng
University of Hong Kong, Hong Kong

Lei Liu
University of Pennsylvania, Philadelphia, PA, USA

Kristine Lund
ICAR Research Laboratory, CNRS, University of Lyon, Lyon, France

Alejandra Martínez-Monés
GSIC-EMIC Group, University of Valladolid, Valladolid, Spain

Richard Medina
Laboratory for Interactive Learning Technologies, Dept. of Information and Computer Sciences, University of Hawai'i, Manoa, USA

Muhsin Menekse
Mary Lou Fulton Institute, Arizona State University, Tempe, AZ, USA

Elena Papanastasiou
University of Nicosia, Nicosia, Cyprus

Carolyn Penstein Rosé
Human–Computer Interaction Institute, Carnegie Mellon University, Pittsburg, PA, USA

Sadhana Puntambekar
University of Wisconsin, Madison, WI, USA

Peter Reimann
The Faculty of Education and Social Work, Education Building A35.252, University of Sydney NSW 2006

Nikol Rummel
Institute of Education, Ruhr-Universität Bochum, Universitätsstraße 150, 44801 Bochum, Germany

Victor D. Sampson
School of Teacher Education, Florida State University, FL, USA

Eileen Scanlon
Institute of Educational Technology, Open University, Milton Keynes, Bucks, MK 7 6AA, UK

Hans Spada
Institute of Psychology, University of Freiburg, Germany

Gerry Stahl
College of Information Science and Technology, Drexel University, 3141 Chestnut St, Philadelphia, USA

Karsten Stegmann
University of Munich, Leopoldstrasse 13, 80802 München, Germany

Daniel Suthers
Laboratory for Interactive Learning Technologies, Department of Information and Computer Sciences, University of Hawai'i, Manoa, HI, USA

Agni Stylianou-Georgiou
Department of Education, University of Nicosia, Nicosia, Cyprus

Stephanie D. Teasley
School of Information, University of Michigan, 4384 North Quad, 105 S. State St., Ann Arbor, MI 48109–1285

John Voiklis
Teachers College, Columbia University, New York, NY, USA

Wing O.W. Wong
University of Hong Kong, Hong Kong

Kalina Yacef
University of Sydney, Australia

Johnny Yuen
University of Hong Kong, Hong Kong

Part I
Understanding Group Processes

Chapter 1
A Complexity-Grounded Model for the Emergence of Convergence in CSCL Groups

Manu Kapur, John Voiklis, and Charles K. Kinzer

Abstract We advance a complexity−grounded, quantitative method for uncovering temporal patterns in CSCL discussions. We focus on convergence because understanding how complex group discussions converge presents a major challenge in CSCL research. From a complex systems perspective, convergence in group discussions is an emergent behavior arising from the transactional interactions between group members. Leveraging the concepts of *emergent simplicity* and *emergent complexity* (Bar-Yam 2003), a set of theoretically-sound yet simple rules was hypothesized: Interactions between group members were conceptualized as goal-seeking adaptations that either help the group move towards or away from its goal, or maintain its status quo. Operationalizing this movement as a Markov walk, we present quantitative and qualitative findings from a study of online problem-solving groups. Findings suggest high (or low) quality contributions have a greater positive (or negative) impact on convergence when they come earlier in a discussion than later. Significantly, convergence analysis was able to predict a group's performance based on what happened in the first 30–40% of its discussion. Findings and their implications for CSCL theory, methodology, and design are discussed.

1.1 Introduction

One of the major challenges facing collaborative problem-solving research is to understand the process of how groups achieve convergence in their discussions (Fischer and Mandl 2005). For example, without some level of convergent or shared assumptions and beliefs collaborators cannot define (perhaps not even recognize) the problem at hand nor select among the possible solutions, much less take action (Schelling 1960). A certain level of convergence—e.g., by what names collaborators refer to the objects of the problem—is required simply to carry

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

S. Puntambekar et al. (eds.), *Analyzing Interactions in CSCL*, Computer-Supported Collaborative Learning Series, DOI 10.1007/978-1-4419-7710-6_1,

on a conversation (Brennan and Clark 1996). A deeper level of convergence—e.g., agreeing on the functional significance of the objects to which partners refer—is required to carry out shared intentions on those objects (Clark and Lucy 1975). Not surprisingly, Roschelle (1996) argued that convergence, as opposed to socio-cognitive conflict, is more significant in explaining why certain group discussions lead to more productive outcomes than others.

Although there has been considerable research towards understanding the cognitive and social mechanisms of convergence in collaborative learning environments (Clark and Brennan 1991; Fischer and Mandl 2005; Jeong and Chi 2007; Roschelle and Teasley 1995; Stahl 2005), the problem of convergence—understanding the complex process of how multiple actors, artifacts, and environments interact and evolve in space and time to converge on an outcome is critical—remains a perennial one (Barab et al. 2001). A substantial amount of literature attempts to understand group processes using qualitative analytical methods (e.g., interactional analysis, discourse analysis, conversation analysis), which provide insightful and meaningful micro-genetic accounts of the complex process of emergence of convergence in groups (e.g., Barron 2000, 2003; Teasley & Roschelle 1993; Stahl 2005). For the present purposes, however, our proposal speaks to quantitative approaches, typically involving quantitative content analysis (QCA) (Chi 1997) of interactional data; the use of QCA is pervasive in examining the nature of interaction and participation in CSCL research (Rourke Anderson 2004). However, quantitative measures and methods for conceptualizing the *temporal* evolution of collaborative problem-solving processes as well as the emergence of convergence remain lacking (Barab et al. 2001; Collazos et al. 2002; Hmelo-Silver et al. this book; Reimann 2009).

Increasingly, a realization of the inherent complexity in the interactional dynamics of group members is giving way to a more emergent view of how groups function and perform (Arrow et al. 2000; Kapur et al. 2007; Kapur et al. 2008; Stahl 2005). However, the use of complex systems in the learning sciences is relatively sparse, but gaining momentum (see Jacobson and Wilensky 2006). A major thrust of such research is on the curricula, teaching, and learnability issues related to complex systems, and how they influence learning and transfer. However, complex systems also offer important theoretical conceptions and methodologies that can potentially expand the research tool-kit in the learning sciences (Jacobson and Wilensky 2006; Kapur and Jacobson 2009; Kapur et al. 2005, 2007). The work reported in this chapter leverages this potential to better understand how convergence emerges in group discussions.

From a complex systems' perspective, convergence in group discussions can be seen as a complex, emergent behavior arising from the local interactions between multiple actors, and mediated by tools and artifacts. Convergence is therefore a group-level property that cannot be reduced to any particular individual in the group. Yet, it emerges from and constrains the interactions between the very individuals it cannot be reduced to. To understand this emergence, we first discuss the concept of emergent behavior, particularly the distinction between *emergent simplicity* and *emergent complexity*; a distinction that is central to our proposal. Following that, we describe one way in which convergence in group discussions can be conceptualized and modeled. We support our case empirically, through findings from a study of CSCL groups. We end by discussing the implications of our work for CSCL theory, methodology, and design.

1.1.1 Unpacking Emergent Behavior: Emergent Simplicity Versus Emergent Complexity

Central to the study of complex systems is how the complexity of a whole is related to the complexity of its parts (Bar-Yam 2003). The concept of emergent behavior—how macro-level behaviors emerge from micro-level interactions of individual agents—is of fundamental importance to understanding this relationship. At the same time, the concept of emergent behavior is rather paradoxical. On the one hand, it arises from the interactions between agents in a system, e.g., individuals in a collective. On the other hand, it constrains subsequent interactions between agents and in so doing, seems to have a life of its own independent of the local interactions (Kauffman 1995), and therefore, cannot be reduced to the very individual agents (or parts) of the system it emerged from (Lemke 2000). For example, social structures (e.g., norms, values, beliefs, lexicons, etc.) within social networks emerge from the local interactions between individual actors. At the same time, these structures shape and constrain the subsequent local interactions between individual actors, but they cannot be reduced to the very individual actors' behaviors and interactions they emerged from (Lemke 2000; Watts and Strogatz 1998). Therefore, it becomes fundamentally important to understand *how* macro-level behaviors emerge from and constrain micro-level interactions of individual agents.

Understanding the "how," however, requires an understanding of two important principles in complexity. First, simple rules at the local level can generate complex emergent behavior at the collective level (Kauffman 1995; Epstein and Axtell 1996). For example, consider the brain as a collection of neurons. These neurons are complex themselves, but exhibit simple binary behavior in their synaptic interactions. This type of emergent behavior, when complexity at the individual-level results in simplicity at the collective-level, is called *emergent simplicity* (Bar-Yam 2003). Further, these simple (binary) synaptic interactions between neurons collectively give rise to complex brain "behaviors"—memory, cognition, etc.—that cannot be seen in the behavior of individual neurons. This type of emergent behavior, when simplicity at the individual-level results in complexity at the collective-level, is called *emergent complexity* (Bar-Yam 2003).

The distinction between emergent simplicity and complexity is critical, for it demonstrates the possibility that a change of scale (micro vs. macro level) can be accompanied with a change in the type (simplicity vs. complexity) of behavior (Kapur and Jacobson 2009). "Rules that govern behavior at one level of analysis (the individual) can cause qualitatively different behavior at higher levels (the group)" (Gureckis and Goldstone 2006, p. 1). We do not necessarily have to seek complex explanations for complex behavior; complex collective behavior may very well be explained from the "bottom up" via simple, minimal information, e.g., utility function, decision rule, or heuristic, contained in local interactions (Kapur et al. 2006; Nowak 2004).

In this chapter, we use notions of emergent simplicity and complexity to conceptualize a group of individuals (agents) interacting with each other as a complex system. The group, as a complex system, consists of complex agents, i.e., just like the neurons, the individuals themselves are complex. Again, it is only intuitive to think that their behavior can be anything but complex and any attempt to model it via simple rules is futile. However, emergent simplicity suggests that this is not an ontological necessity. Their behavior may very well be modeled via simple rules. Further, emergent complexity suggests that doing so may reveal critical insights into the complexity of their behavior as a collective. It is this possibility that we explore and develop in the remainder of this chapter.

1.1.2 Purpose

We describe how convergence in group discussions can be examined as an emergent behavior arising from theoretically-sound yet *simple* teleological rules to model the collaborative, problem-solving interactions of its members. We support our model empirically, through findings from a study of groups solving problems in an online, synchronous, chat environment. Note that this study was part of a larger program of research on *productive failure* (for more details, see Kapur 2008, 2009, 2010; Kapur and Kinzer 2007, 2009). We first describe the context of the study in which the methodology was instantiated before illustrating the methodology.

1.2 Methodology

1.2.1 Research Context and Data Collection

Participants included sixty 11th-grade students (46 male, 14 female; 16–17 years old) from the science stream of a co-educational, English-medium high school in Ghaziabad, India. They were randomized into 20 triads and instructed to collaborate and solve either well- or ill-structured problem scenarios. The study was carried out in the school's computer laboratory, where group members communicated with one another only through synchronous, text-only chat. The 20 automatically-archived transcripts, one for each group, contained the group discussions as well as their solutions, and formed the data used in our analyses.

1.2.2 Procedure

A well-structured (WS) and an ill-structured (IS) problem scenario were developed consistent with Jonassen's (2000) design theory typology for problems (see Kapur 2008 for the problem scenarios). Both problem scenarios dealt with a car accident scenario and targeted the same concepts from Newtonian Kinematics and Laws of Friction to solve them. Content validation of the two problem scenarios was achieved with the help of two physics teachers from the school with experience in teaching those subjects at the senior secondary levels. The teachers also assessed the time students needed to solve the problems. Pilot tests with groups of students from the previous cohort further informed the time allocation for the group work, which was set at 1.5 h.

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

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

1.2.3 Hypothesizing Simple Rules

The concept of *emergent simplicity* was invoked to hypothesize a set of simple rules. Despite the complexity of the individual group members, the impact of their interactions was conceived to be governed by a set of simple rules. Group members were conceived as agents interacting with one another in a goal-directed manner toward solving a problem. Viewed *a posteriori*, these transactional interactions seemed to perform a telic function, i.e., they operated to reduce the difference between the current problematic state of the discussion and a goal state. Thus, local interactions between group members can be viewed as operators performing a means-ends analysis in the problem space (Newell and Simon 1972). From this, a set of simple rules follows naturally. Each interaction has an impact that:

1. Moves the group *towards* a goal state, or
2. Moves the group *away* from a goal state, or
3. *Maintains* the status-quo (conceptualized as a "neutral impact").

Then, convergence in group discussion was conceived as an *emergent complexity* arising from this simple-rule-based mechanism governing the impact of individual agent-based interactions.

1.2.4 Operationalizing Convergence

Concepts from the statistical theory of Markov walks were employed to operationalize the model for convergence (Ross 1996). Markov walks are commonly used to model a wide variety of complex phenomenon (Ross 1996). First, quantitative content analysis (QCA; Chi 1997), also commonly known as coding and counting, was used to segment utterances into one or more interaction units. The interaction unit of analysis was semantically defined as the impact(s) that an utterance had on the group discussion vis-à-vis the hypothesized simple rule. Two trained doctoral students independently coded the interactions with an inter-rater reliability (*Krippendorff's alpha*) of.85. An impact value of 1, −1, or 0 was assigned to each interaction unit depending upon whether it moved the group discussion toward (impact = 1) or away (impact = −1) from the goal of the activity—a solution state of the given problem—or maintained the status quo (impact = 0). Therefore, each discussion was reduced to a temporal string of 1 s, −1 s, and 0 s.

More formally, let n_1, n_{-1}, and n_0 denote the number of interaction units assigned the impact values 1, −1, and 0 respectively up to a certain utterance in a discussion. Then, up to that utterance, the convergence value was,

$$C = \frac{n_1 - n_{-1}}{n_1 + n_{-1}}.$$

For each of the 20 discussions, convergence values were calculated after each utterance in the discussion, resulting in a notional time series representing the evolution of convergence in group discussion.

1.3 Results

Plotting the convergence value on the vertical axis and time (defined notionally with utterances as *ticks* on the evolutionary clock) on the horizontal axis, one gets a representation (also called a fitness curve) of the problem-solving process as it evolves

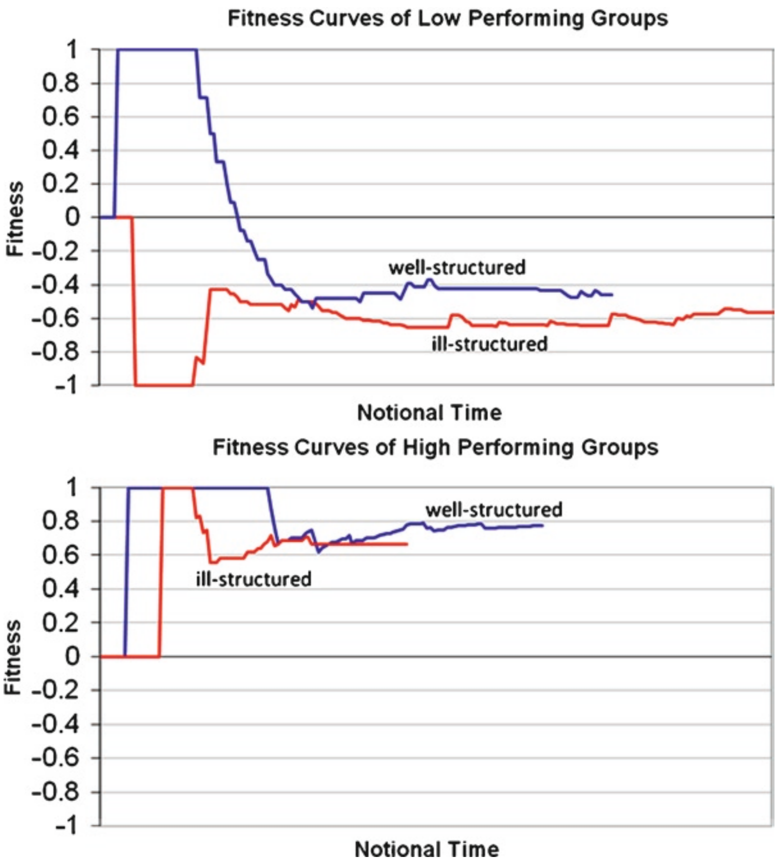

Fig 1.1 Fitness curves of high and low performing groups across problem types

in time. Figure 1.1 presents four major types of fitness curves that emerged from the discussion of the 20 problem-solving groups in our study. These four fitness curves contrast the high- with the low-performing groups (group performance is operationalized in the next section) across the well- and ill-structured problem types.

1.3.1 *Interpreting Fitness Curves*

It is easy to see that the convergence value always lies between –1 and 1. The closer the value is to 1, the higher the convergence, and the closer the group is to reaching a solution. The end-point of the fitness curve represents the final fitness level or convergence of the entire discussion. From this, the extent to which a group was successful in solving the problem can be deduced. Furthermore, one might imagine that an ideal fitness curve is one that has all the moves or steps in the positive direction, i.e., a horizontal straight line with fitness equaling one. However, the data suggests that, in reality, some level of divergence of ideas may in fact be a good thing (Kapur 2008, 2009; Schultz-Hardt et al. 2002), as can be seen in the fitness curves of both the high-performing groups.

The shape of the fitness curve, therefore, is also informative about the paths respective groups take toward problem solution. For example, in Fig. 1.1, both the low-performing groups converged at approximately the same (negative) fitness levels, but their paths leading up to their final levels were quite different. The well-structured group showed a sharp fall after initially moving in the correct direction (indicated by high fitness initially). The ill-structured group, on the other hand, tried to recover from an initial drop in fitness but was unsuccessful, ending up at approximately the same fitness level as the well-structured group. Further, comparing the high-performing with the low-performing groups, one can see that the discussions of high-performing groups had fewer utterances, regardless of problem type. Finally, all fitness curves seemed to settle into a *fitness plateau* fairly quickly. What is most interesting is that this descriptive examination of fitness curves provides a view of paths to a solution that are lost in analysis systems that consider only a given point in the solution process, thus assuming that similar behaviors or states at a given point are arrived at in similar ways. As different paths can lead to similar results, unidimensional analyses that consider only single points in time (often only the solution state) are not consistent with what this study's data suggest about problem-solving processes.

Most important is a mathematical property of convergence. Being a ratio, convergence is more sensitive to initial contributions, both positive and negative, than those made later in the process. This can be easily seen because with each positive (or negative) contribution, the ratio's numerator is increased (or decreased) by one. However, the denominator in the ratio always increases, regardless of the contribution being positive or negative. Therefore, when a positive (or negative) contribution comes earlier in the discussion, its impact on convergence is greater because a unit increment (or decrement) in the numerator is accompanied by a denominator that is smaller earlier than later. Said another way, this conceptualization of convergence allows us to test the following hypothesis: *"good" contributions made earlier in a group discussion, on average, do more good than if they were made later. Similarly, "bad" ones, on average, do more harm if they come earlier than later in the discussion*. To test this hypothesis, the relationship between convergence and group performance was explored by running a temporal simulation on the data set.

1.3.2 Relationship between Convergence and Group Performance

The purpose of the simulation was to determine if the level of convergence in group discussion provided an indication of the eventual group performance. Group performance was operationalized as the quality of group solution, independently rated by two doctoral students on a nine-point rating scale (Table 1.1) with an inter-rater reliability (*Krippendorff's alpha*) of .95.

Table 1.1 Rubric for coding quality of group solution

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

Mid-point scores of .5, 1.5, 2.5, and 3.5 were assigned when the quality of solution was assessed to be between the major units 0, 1, 2, 3, and 4, making the scale essentially a 9-point scale

The discussions of all 20 groups were each segmented into ten equal parts. At each tenth, the convergence value up to that point was calculated. This resulted in 10 sets of 20 convergence values; the first set corresponding to convergence in the discussion after 10% of the discussion was over, the second after 20% of the discussion was over, and so on until the tenth set, which corresponded to the final convergence value of the discussion, i.e., after 100% of the discussion had occurred. A simulation was then carried out by regressing group performance on convergence values at each tenth of the discussion (hence, a temporal simulation), controlling for problem type (well- or ill-structured) each time. The *p*-value corresponding to the statistical significance of the predictive power of convergence at each tenth of the discussion on eventual group performance was plotted on the vertical axis (see Fig. 1.2).

C1 through C10 denote the ten equally spaced instances in each discussion at which the convergence values were calculated. The simulation suggested that, on average, at some point after 30% but before 40% of the discussion is over (i.e., between C3 and C4 in Fig. 1.2), the convergence value is able to predict eventual group performance at the.05 level of significance or better. This shows that convergence is a powerful measure that is able to model the impact that early contributions have on eventual group performance. This insight bears important implications for scaffolding group discussions to achieve optimal outcomes. For example, if one's primary aim is to maximize group performance, the insight suggests a need for scaffolding early in the discussion, since the impact of early interactional activity on eventual group performance seems to be greater. Scaffolding earlier parts of a group discussion may increase its likelihood of settling into higher fitness plateaus; the higher the fitness plateau, the better the group performance, on average. This insight is in and of itself a significant finding, but since participation in high-performing groups is consistently (and not surprisingly) a strong predictor of subsequent individual learning gains (e.g., see Barron 2003; Cohen et al. 2002), we believe it makes it all the more significant, for it demonstrates strong connections to group and individual learning.

To delve deeper into what makes convergence a powerful measure, micro-analytical interactional analysis sheds more light. We present a brief analysis of the following excerpt containing an exchange between group members S1 and S2.

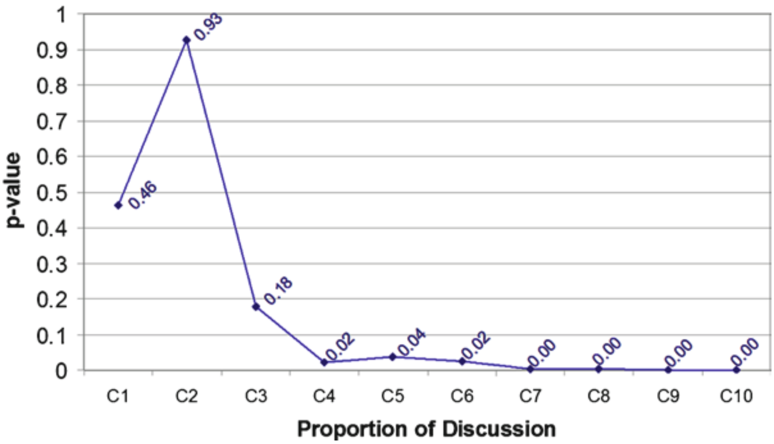

Fig. 1.2 Simulation of the significance of convergence in predicting group performance

S1> *are we going to apply frictional retardation for the reaction time also?*	−1
S2> *no, because reaction time is the time between watching the boy and applying the brakes so in this time [the] car must be accelerating*	1, 1
S1> *but I think we must not forget that the car is moving on the same road on which the incident occurs and the road is providing the retardation*	−1, −1
S2> *but maximum speed is at the instant when he applied the brake*	1
S1> *but earlier you said that the car will accelerate after perceiving the boy*	−1
S2> *I said so because his foot must be on accelerator during reaction time*	1
S1> *Now I understand… please proceed to write your answer*	1, 1

Recall that the problem involved a car-accident scenario (see Kapur 2008, for the problem scenario). In this excerpt, S1 and S2 are trying to decide whether or not reaction time of the driver of the car that was involved in the accident should factor into their calculations. The excerpt starts with S1 asking a question about applying frictional retardation during the driver's reaction time. Being a misconception, it was rated as having a negative impact (−1). S2 evaluates S1's question and says '*no*,' attempting to correct the misconception. Hence, its positive (+1) impact rating. In the same utterance, S2 elaborates why frictional retardation should not to be applied, further positively impacting the group's progress. The argument continues with S1 persisting with the misconception (assigned negative impacts) until S2 is able to convince S1 otherwise (assigned positive impacts), thereby converging on a correct understanding of this aspect (dealing with friction during reaction time) of the problem. Note that had S2 wrongly evaluated and agreed to S1's misconception, the impact ratings would have been negative, which, without any further correction, would have led the group to diverge from a correct understanding of that very aspect of the problem.

This analysis, albeit brief, shows that impact ratings are meaningful only in relation to preceding utterances (Bransford and Nitsch 1978) and take into account the sequence and temporality of collaborative interactions (Kapur et al. 2008). Other examples of highly convergent discussion episodes would include agreement with and positive evaluation and development of correct understandings of the problem, solution proposals, and problem solving strategies. As a result, despite solving different types of problems (well- or ill-structured), group performance depended mainly upon the convergence of their discussions. Because convergence takes into account both the number as well as the temporal order of the units of analyses, it utilizes a greater amount of information present in the data. This makes convergence a more powerful measure, both conceptually and statistically, than existing predictors that do not fully utilize the information present in interactional data. If this is the case, then the following hypothesis should hold: *convergence is a more powerful predictor of group performance than existing, commonly-used interactional predictors*.

1.3.3 Comparing Convergence with Other Commonly-Used Interactional Predictors

Many studies of collaborative problem solving, including this one, use QCA to operationalize measures for problem-solving processes. These measures typically result in data about the frequency or relative frequency of positive indicators (e.g., higher-order thinking, questioning, reflecting, etc.), or negative indicators (e.g., errors, misconceptions, lack of cooperation, non-task, etc.), or a combination that adjusts for both the positive and negative indicators (e.g., the difference between the frequencies of high- and low-quality contributions in a discussion). In this study, we operationalized three measures to represent typical measures:

1. Frequency (n_1: recall that this is the number of interaction units in a discussion with impact = 1),
2. Relative Frequency, $\left(\frac{n_1}{n_1 + n_0 + n_{-1}}\right)$, and
3. Position, $(n_1 - n_{-1})$.

Convergence, $\left(\frac{n_1 - n_{-1}}{n_1 + n_{-1}}\right)$, formed the fourth measure.

Multiple linear regression was used to simultaneously compare the significance of the four measures in predicting group performance, controlling for problem type in each case. The overall model was significant, $F = 6.391$, $p = .003$. Results in Table 1.2 suggest that, of the four predictors of group performance, convergence was the only one significant, $t = 2.650$, $p = .019$, thereby supporting our hypothesis. In other words, consistent with our hypothesis, convergence seems to be a more powerful predictor of problem-solving performance when compared to existing, commonly-used predictors.

Table 1.2 Regression parameter estimates of the interactional variables and problem type

	B	SE	Beta	t	p
(Constant)	−1.778	1.693		−1.051	.311
Frequency	.006	.021	.239	.268	.793
Relative frequency	1.541	2.792	.116	.552	.590
Position	−.012	.024	−.446	−.513	.616
Convergence	5.338	2.014	.839	2.650	.019
Problem type	−.050	.544	−.021	−.092	.928

1.4 Discussion

In this chapter, we have described a complexity-grounded model for convergence. We argued that convergence in group discussions can be seen as a complex, emergent behavior arising from the local interactions between multiple actors, and mediated by tools and artifacts. That is, convergence is a group-level property that cannot be reduced to any particular individual in the group. Yet, it emerges from and constrains the interactions between the very individuals it cannot be reduced to. A complexity-grounded model allowed us to model a complex, group-level emergent behavior such as convergence using simple interactional rules between group members. More specifically, we used the concepts of emergent simplicity and emergent complexity to hypothesize a set of theoretically-sound yet simple rules to model the problem-solving interactions between group members, and then examined the resulting emergent complexity: Convergence in their discussion.

Despite the intentional simplicity of our model, it revealed novel insights into the process of collaboration. The first insight concerned the differential impact of contributions in a group discussion—high (or low) quality contributions have a greater positive (or negative) impact on the eventual outcome when they come earlier than later in a discussion. A corollary of this finding was that eventual group performance could be predicted based on what happens in the first 30–40% of a discussion because group discussions tended to settle into fitness plateaus fairly quickly. Finally, convergence was shown to be a more powerful predictor of group performance than some existing, commonly-used measures. These insights are significant, especially since participation in high-performing groups is a strong predictor of subsequent individual learning gains (e.g., see Barron 2003; Cohen et al. 2002). In other words, this conceptualization and analysis of convergence demonstrates strong connections to both group performance and individual learning.

1.4.1 Implications for Scaffolding

If, as our work suggests, group performance is highly sensitive to early exchanges in a discussion, then this insight bears important implications for scaffolding synchronous, small-group, CSCL discussions to achieve optimal outcomes. For example, if one primarily aims to maximize group performance, these early sensitivities suggest a need for scaffolding early in the discussion, since the impact of early interactional activity on eventual group performance seems to be greater. Scaffolding earlier parts of a group discussion may increase its likelihood of settling into higher fitness plateaus; better group performance is predicated by high fitness plateaus. This is also consistent with the notion of fading, that is, having scaffolded the early exchange, the scaffolds can be faded (Wood et al. 1976). For example, instead of scaffolding the entire process of problem solving using process scaffolds, it may only be necessary to scaffold how a group analyzes and frames the problem, as these problem categorization processes often occur early in problem-solving discussions and can shape all subsequent processes (Kapur and Kinzer 2007; Kapur et al. 2005; Voiklis 2008). Such an approach stands in contrast with the practice of blanket scaffolding of the CSCL processes (e.g., through the use of collaborative scripts). The above are testable hypotheses that emerge from this study and we invite the field to test and extend this line of inquiry.

1.4.2 Implications for Methodology: The Temporal Homogeneity Assumption

Sensitivity to early exchange also underscores the role of temporality, and consequently, the need for analytical methods to take temporal information into account. According to Reimann (2009), "Temporality does not only come into play in quantitative terms (e.g., durations, rates of change), but order matters: Because human learning is inherently cumulative, the sequence in which experiences are encountered affects how one learns and what one learns" (p. 1). Therefore, understanding (1) how processes evolve in time, and (2) how variation in this evolution explains learning outcomes, ranks among the more important challenges facing CSCL research (Akhras and Self 2000; Hmelo-Silver et al. this book; Reimann 2009).

To derive methodological implications, let us first consider a prototypical case of coding and counting in CSCL. Typically, one or more coding/rating schemes are applied to the interaction data, resulting in a cumulative frequency or relative frequency distribution of interactions across the categories of the coding/rating scheme (e.g., depth of explanations, functional content of interactions, misconceptions, quality, correctness, etc.). These distributions essentially tally the amount, proportion, and type of interactions vis-à-vis the interactional coding/rating scheme (Suthers 2006). Significant links are then sought between quantitatively-coded interactional data and outcomes, such as quality of group performance and group-to-individual transfer (see Rourke and Anderson 2004, for a discussion on the validity of QCA).

Notwithstanding the empirically-supported significant links between the nature of interactional activity and group performance, interpreting findings from interactional coding/rating schemes is limited by the very nature of the information tapped by these measures. For example, what does it mean if a group discussion has a high proportion of a certain category of interaction? It could be that interactions coded in that category were spread throughout the discussion, or perhaps they were clustered together in a coherent phase during the discussion. Therefore, interactions that are temporally far apart in a discussion carry the same weight in the cumulative count or proportion: one that comes later in a discussion is given the same weight as one that comes earlier. Such an analysis, while informative, does not take the temporality of interactions into account, i.e., the time order of interactions in the problem-solving process. By aggregating category counts over time, one implicitly makes the assumption of *temporal homogeneity* (Kapur et al. 2008). In light of the complexities of interactional dynamics in CSCL, it is surprising how frequently this assumption of temporal homogeneity is made without justification or validation (Voiklis et al. 2006).

It follows, then, that we need methods and measures that take temporality into account. These methods and measures can potentially allow us to uncover patterns in time and reveal novel insights (e.g., sensitivity to early exchange) that may otherwise not be possible. Consequently, these methods and measures can play an instrumental role in the building and testing of a process-oriented theory of problem solving and learning (Reimann 2009).

1.4.3 Implications for Theorizing CSCL Groups as Complex Systems

Interestingly, sensitivity to early exchange exhibited by CSCL groups in our study seems analogous to sensitivity to initial conditions exhibited by many complex adaptive systems (Arrow et al. 2000; Bar-Yam 2003); the idea being that small changes initially can lead to vastly different outcomes over time, which is what we found in our study. Furthermore, the locking-in mechanism is analogous to *attractors* in the phase space of complex systems (Bar-Yam 2003). Phase space refers to the maximal set of states a complex system can possibly find itself in as it evolves. Evidently, a group discussion has an infinite phase space, yet the nature of early exchange can potentially determine whether it organizes into higher or lower fitness attractors. Thus, an important theoretical and methodological implication from this finding is that CSCL research needs to pay particular attention to the temporal aspects of interactional dynamics (Hmelo-Silver et al. this book). As this study demonstrates, studying the evolution of interactional patterns can be insightful, presenting counterintuitive departures from assumptions of linearity in, and temporal homogeneity of, the problem solving process (Voiklis et al. 2006).

At a more conceptual level, the idea that one can derive meaningful insights into a complex interactional process via a simple rule-based mechanism, while compelling, may also be unsettling and counter-intuitive. Hence, a fair amount of intuitive resistance to the idea is to be expected. For instance, it is reasonable to argue that the extreme complexity of group interaction—an interweaving of syntactic, semantic, and pragmatic structures and meanings operating at multiple levels—make it a different form of emergence altogether and, therefore, insights into complex interactional processes cannot be gained by using simple-rule-governed methods. However, a careful consideration of this argument reveals an underlying ontological assumption that complex behavior cannot possibly be explained by simple mechanisms. Saying it another way, some may argue that *only* complex mechanisms (e.g., linguistic mechanisms) can explain complex behavior (e.g., convergence in group discussion). Of course, this is a possibility, but, notions of emergent simplicity and emergent complexity suggest that this is not the "only" possibility (Bar-Yam 2003), especially given our knowledge of the laws of self-organization and complexity (Kauffman 1995).

It is noteworthy that emergent complexity is also integral to the theory of dynamical minimalism (Nowak 2004) used to explain complex psychological and social phenomena. Dynamical minimalist theory attempts to reconcile the scientific principle of parsimony—that simple explanations are preferable to complex ones in explaining a phenomenon—with the arguable loss in depth of understanding of that phenomenon because of parsimony. Using the principle of parsimony, the theory seeks the simplest mechanisms and the fewest variables to explain a complex phenomenon. It argues that this need not sacrifice depth in understanding because simple rules and mechanisms that repetitively and dynamically interact with each other can produce complex behavior: the very definition of emergent complexity. Thus, parsimony and complexity are not irreconcilable, leading one to question the assumption that complex phenomena necessarily require complex explanations (Kapur and Jacobson 2009).

Therefore, the conceptual and methodological implication from this study is not that complex group behavior ought to be studied using simple-rule-based mechanisms, but that exploring the possibility of modeling complex group behavior using simple rule-based mechanisms is a promising and meaningful endeavor. Leveraging this possibility, we demonstrated one way in which simple-rule-based mechanisms can be used to model convergence in group discussion, in turn revealing novel insights into the collaborative process. The proposed measures of convergence and fitness curves were intentionally conceived and designed to be generic and, therefore, may be potentially applicable to other problem-solving situations as well. Thus, they also provide a platform for the development of more sophisticated measures and techniques in the future.

1.4.4 Some Caveats and Limitations

New methods and measures always raise more questions than answers, and ours is not an exception. What is more important is that repeated application and modification over multiple data sets is needed before strong and valid inferences can be made (Rourke and Anderson 2004). At this stage, therefore, our findings remain technically bound by the context of this study; it is much too early to attempt any generalization. There are also several issues that need to be highlighted when considering the use of the proposed methodology:

Issues of coding: Clearly, drawing reliable and valid inferences based on the new measures minimally requires that the coding scheme be reliable and valid. To this end, a conscious, critical decision was our choice of the content domain: we chose Newtonian kinematics because it is a relatively well-structured domain. This domain structure clarified the task of differentiating those contributions to the problem-solving discussion that moved the group closer to a solution from those that did not and, thus, minimized the effort of coding the impact of interactional units (1, −1, and 0). For a more complex domain (e.g., ethical dilemmas) where the impact is not as easy to assess, our method may not be as reliable, or perhaps not even applicable.

Model simplicity: It can be argued that the proposed model is a very simple one, and could be seen as a limitation. The decision to keep the model simple was intentional; we chose to keep the number of codes to a minimum, i.e., just −1, 0, or 1. We reasoned that if we could reveal novel insights by using the simplest model, then one could always "complexify" the model subsequently. For example, one could easily build on this model to code impact on a five point scale from −2 to 2 so as to discriminate contributions that make a greater positive or negative impact than others. At the same time, the model also collapses many dimensions (such as social, affective, cognitive, meta-cognitive, and so on) into one dimension of impact. Collapsing dimensions into a simple model allows for the easy and direct interpretation of results, but this gain in interpretability comes with the cost of an overly reductive model. Once again, we wanted to demonstrate that even with a simple model, one could potentially gain insights, and having done so, one could always embark on a building a more complex model. For example, it might be useful to model the co-evolution of the various dimensions, investigate the co-evolving fitness trajectories, and develop deeper understandings of the phenomenon.

Corroborating interpretations: Our model is essentially a quantitative model. In reducing complex interactions into impact ratings, it is necessarily reductive. In interpreting findings from such analyses, it is important to use the method as part of a mixed-method analytical commitment. If not, it may be hard to differentiate results that are merely a statistical or mathematical artifact from the ones that are substantively and theoretically meaningful.

1.4.5 Future Directions

Going forward, we see the need for developing new temporal measures. We want to focus particularly on those that can be easily implemented from a temporal sequence of codes that QCA of group discussions normally results in. In particular, we argue for two candidates:

1. **Lag-Sequential Analysis (LSA):** LSA treats each interactional unit (as defined in a study) as an observation; a coded sequence of these observations forming the

interactional sequence of a group discussion (e.g., Erkens et al. 2003). LSA detects the various non-random aspects of interactional sequences to reveal how certain types of interactions follow others more often than what one would expect by chance (Wampold 1992). By examining the transition probabilities between interactions, LSA identifies statistically significant transitions from one type of interactional activity to another (Bakeman and Gottman 1997; Wampold 1992). As a result, the collaborative process can be examined as an evolving, multi-state network, thereby allowing us to reveal temporal patterns that may otherwise remain hidden (Kauffman 1995). For example, Kapur (2008) coded collaborative problem solving interactions into process categories of problem analysis, problem critique, criteria development, solution development, and solution evaluation, thereby reducing each group discussion into a temporal string of process category events. Using LSA, the analysis revealed significant temporal patterns that the typical coding and counting method could not reveal, that is, how some process categories were more likely to follow or be followed by other process categories significantly above chance level. More importantly, LSA demonstrated how variation in temporal patterns—sequences of process categories—was significantly related to variation in group performance.

2. **Hidden Markov Models (HMMs): HMMs** (Rabiner 1989) offer analysis at a relatively coarser grain size than LSA by detecting the broader interaction phases that a discussion goes through. For example, Soller and colleagues (2002) used HMMs to analyze and assess temporal patterns in on-line knowledge sharing conversations over time. Their HMM model could determine the effectiveness of knowledge sharing phases with 93% accuracy, that is, 43% above what one would expect by chance. They argued that understanding these temporal phases that provide an insight into the dynamics of how groups share, assimilate, and build knowledge together is important in building a process theory of facilitating to increase the effectiveness of the group interactions. Conceiving a group discussion as a temporal sequence of phases, one can use several methods to isolate evolutionary phases, including measures of genetic entropy (Adami et al. 2000), intensity of mutation rates (Burtsev 2003) or, in the case of problem interactions, the classification of coherent phases of interaction. With the phases identified, one can calculate and predict the probabilities between phases using HMMs (Rabiner 1989; for an example, see Holmes 1997). As a result, one may begin to understand when and why phase transitions as well as stable phases emerge; more importantly, one may begin to understand how the configuration of one phase may influence the likelihood of moving to any other phase. Whether one can control or temper these phases, or whether such control or temperance would prove a wise practice remains an open question which, even if only partially answered, will be a breakthrough in characterizing and modeling the problem solving process.

It is worth reiterating that these methods should not be used in isolation, but as part of a larger, multi-method, multiple grain size analytical program. At each grain size, findings should potentially inform and be informed by findings from analysis at other grain sizes—an analytical approach that is commensurable with the multiple levels (individual, interactional, group) at which the phenomenon unfolds. Only then can these methods and measures can play an instrumental role in the building and testing of a process-oriented theory of problem solving and learning (Hmelo-Silver et al. this book; Reimann 2009; Reimann et al. this book).

1.5 Conclusion

In this chapter, we have advanced a complexity-grounded, quantitative method for uncovering temporal patterns in interactional data from CSCL discussions. In particular, we have described how convergence in group discussions can be examined as an emergent behavior arising from theoretically-sound yet simple teleological rules to model the collaborative, problem-solving interactions of its members. We were able to design a relatively simple model to reveal a preliminary yet compelling insight into the nature and dynamics of problem-solving CSCL groups. That is, convergence in group discussion, and consequently group performance, is highly sensitive to early exchanges in the discussion. More importantly, in taking these essential steps toward understanding of how temporality affects CSCL group processes and performance, we call for further efforts within this line of inquiry.

Acknowledgments The work reported herein was funded by a Spencer Dissertation Research Training Grant from Teachers College, Columbia University to the first author. This chapter reports work that has, in parts, been presented at the International Conference of the Learning Sciences in 2006, and the Computer-Supported Collaborative Learning Conference in 2007. Our special thanks go to the students and teachers who participated in this project. We also thank June Lee and Lee Huey Woon for their help with editing and formatting.

References

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

Akhras, F. N., & Self, J. A. (2000). Modeling the process, not the product, of learning. In S. P. Lajoie (Ed.), *Computers as cognitive tools* (No more walls, Vol. 2, pp. 3–28). Mahwah: Erlbaum.

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

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

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

Barron, B. (2000). Achieving coordination in collaborative problem-solving groups. *Journal of the Learning Sciences, 9*, 403–436.

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

Bar-Yam, Y. (2003). *Dynamics of complex systems*. New York: Perseus.

Bransford, J. D., & Nitsch, K. E. (1978). Coming to understand things we could not previously understand. In J. F. Kavanaugh & W. Strange (Eds.), *Speech and language in the laboratory, school, and clinic* (pp. 267–307). Harvard: MIT Press.

Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*(6), 1482–1493.

Burtsev, M. S. (2003). Measuring the dynamics of artificial evolution. In W. Banzhaf, T. Christaller, P. Dittrich, J. T. Kim, & J. Ziegler (Eds.), *Advances in artificial life. Proceedings of the 7th European conference on artificial life* (pp. 580–587). Berlin: Springer.

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

Clark, H., & Brennan, S. (1991). Grounding in communication. In L. B. Resnick, J. Levine, & S. D. Teasley (Eds.), *Perspectives on socially shared cognition* (pp. 127–149). Washington, DC: APA.

Clark, H. H., & Lucy, P. (1975). Understanding what is meant from what is said: A study in conversationally conveyed requests. *Journal of Verbal Learning and Verbal Behavior, 14*(1), 56–72.

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

Collazos, C., Guerrero, L., Pino, J., & Ochoa, S. (2002). *Evaluating collaborative learning processes. Proceedings of the 8th international workshop on groupware (CRIWG'2002)* (pp. 203–221). Berlin: Springer.

Epstein, J. M., & Axtell, R. (1996). *Growing artificial societies: Social science from the bottom up*. Washington, DC/Harvard: Brookings Institution Press/MIT Press.

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

Fischer, F., & Mandl, H. (2005). Knowledge convergence in computer-supported collaborative learning: The role of external representation tools. *Journal of the Learning Sciences, 14*(3), 405–441.

Gureckis, T. M., & Goldstone, R. L. (2006). Thinking in groups. *Pragmatics and Cognition, 14*(2), 293–311.

Hmelo-Silver, C. E., Jordan, R., Liu, L., & Chernobilsky, E. (this book). Representational tools for understanding complex computer-supported collaborative learning environments. In S. Puntambekar, G. Erkens, & C. Hmelo-Silver (Eds.), *Analyzing interactions in CSCL: Methods, approaches and issues* (pp. 83–106) Springer.

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

Jacobson, M. J., & Wilensky, U. (2006). Complex systems in education: Scientific and educational importance and implications for the learning sciences. *Journal of the Learning Sciences, 15*(1), 11–34.

Jeong, H., & Chi, M. T. H. (2007). Knowledge convergence during collaborative learning. *Instructional Science, 35*, 287–315.

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

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

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

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

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

Kapur, M., & Jacobson, M. J. (2009). Learning as an emergent phenomenon: Methodological implications. Paper presented at the annual meeting of the American educational research association. San Diego.

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

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

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

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

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

Kauffman, S. (1995). *At home in the universe: The search for the laws of self-organization and complexity*. New York: Oxford University Press.

Lemke, J. L. (2000). Across the scales of time: Artifacts, activities, and meanings in ecosocial systems. *Mind, Culture and Activity, 7*(4), 273–290.

Newell, A., & Simon, H. (1972). *Human problem solving*. Englewood Cliffs: Prentice Hall.

Nowak, A. (2004). Dynamical minimalism: Why less is more in psychology. *Personality and Social Psychology Review, 8*(2), 183–192.

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

Reimann, P., Yacef, K., & Kay, J. (this book). Analyzing collaborative interactions with data mining methods for the benefit of learning. In S. Puntambekar, G. Erkens, & C. Hmelo-Silver (Eds.), *Analyzing interactions in CSCL: Methods, approaches and issues* (pp. 161–185). Springer.

Roschelle, J. (1996). Learning by collaborating: Convergent conceptual change. In T. Koschmann (Ed.), *CSCL: Theory and practice of an emerging aradigm* (pp. 209–248). Mahwah: Erlbaum.

Roschelle, J., & Teasley, S. D. (1995). The construction of shared knowledge in collaborative problem solving. In C. E. O'Malley (Ed.), *Computer-supported collaborative learning* (pp. 69–197). Berlin: Springer.

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

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

Schelling, T. C. (1960). *The strategy of conflict*. Cambridge: Harvard University Press.

Schultz-Hardt, S., Jochims, M., & Frey, D. (2002). Productive conflict in group decision making: Genuine and contrived dissent as strategies to counteract biased information seeking. *Organizational Behavior and Human Decision Processes, 88*, 563–586.

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

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

Teasley, S. D., & Roschelle, J. (1993). Constructing a joint problem space: The computer as a tool for sharing knowledge. In S. P. Lajoie & S. D. Derry (Eds.), *Computers as Cognitive Tools* (pp. 229–258). Hillsdale, NJ: Erlbaum.

Voiklis, J. (2008). A thing is what we say it is: Referential communication and indirect category learning. PhD thesis, Columbia University, New York.

Voiklis, J., Kapur, M., Kinzer, C., & Black, J. (2006). An emergentist account of collective cognition in collaborative problem solving. In R. Sun (Ed.), *Proceedings of the 28th annual conference of the cognitive science society* (pp. 858–863). Mahwah: Erlbaum.

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

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of "small-world" networks. *Nature, 393*, 440–442.

Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 17*, 89–100.

Chapter 2
Analyzing the Contextual Nature of Collaborative Activity

Maarit Arvaja

Abstract This chapter discusses a methodology designed to explore the contextual nature of collaborative activity. The methods that can be generally considered to be based on '*socio-cultural' discourse analysis* are discussed as a means to explore how different aspects of a situation mediate students' shared meaning-making. First, an analysis is demonstrated, illustrating how different immediate and mediated contexts are embedded in students' discourse as they are engaged in face-to-face collaborative activity in a computer-mediated context. Second, a multidimensional coding scheme is presented for analyzing the contextualized process of collaborative knowledge construction in an asynchronous web-based discussion. The strengths and weaknesses of the methods are also discussed.

2.1 Taking into Account the Context of Activity in Research on Collaboration: Theoretical Considerations

In the area of Computer-Supported Collaborative Learning (CSCL) research, interest has increasingly shifted from the outcomes and products of collaborative work towards the analysis of the processes of collaboration. This shift shows an attempt to gain understanding about the nature of productive joint activity and to identify interactional features that are important for collaborative learning. There is, for example, a wide field of research on studying the features of argumentative interaction and how it can be best supported in different learning contexts (e.g., Baker et al. 2007). The studies focusing on productive discussion have approached the processes of collaboration mostly from a cognitive and social perspective. The quality of collaboration has been studied by examining the amount and types of

M. Arvaja (✉)
Institute for Educational Research, University of Jyväskylä, P.O. Box 35, Jyväskylä 40014, Finland
e-mail: maarit.arvaja@jyu.fi

S. Puntambekar et al. (eds.), *Analyzing Interactions in CSCL*, Computer-Supported Collaborative Learning Series, DOI 10.1007/978-1-4419-7710-6_2,

productive talk that occur. Typically, the categories of talk are defined at the level of utterance, for example by analyzing communicative or strategic functions of utterances (e.g., Fischer et al. 2002), or at the level of meaningful sequences of utterances, for example, by analyzing the cognitive level of episodes (e.g., van Boxtel et al. 2000). However, the weakness of the studies focusing just on the functions or the cognitive level of interaction is that they reveal little about the process of collaboration, as this kind of analysis discards the content and nature of knowledge construction that takes place in interaction between collaborators as well as the context of their activity (Stahl 2002; Crook 1999).

According to Wertsch (1991), however, it is not possible to study thinking and cognition independently from the social, cultural, and historical settings in which they occur. Cognition is a social process embedded within a historically shaped material world, in the sense that it relies on conceptual and material tools that originate from the culture (Bliss and Säljö 1999). To understand collaborative activity and learning, we need to make sense of the discourse students engage in and the tools that mediate their learning (Hmelo-Silver 2003). This means that in order to analyze collaborative learning, we need to go beyond analyzing structures of talk separated from their contexts to also explore how physical and socio-cultural aspects are manifested in students' discursive activity (Black 2007).

The theoretical background of the methodology presented in this chapter leans on the socio-cultural approach to learning (Vygotsky 1978; Wertsch 1991). Even though this theoretical perspective provides an understanding of how learning is socio-culturally situated and how tools - both physical and psychological – mediate learning, we need specified analytical tools for evaluating collaborative activity and learning embedded in specific contexts. The overall analytical approach presented in this chapter can be regarded as based on 'socio-cultural' discourse analysis, as the methods aim to explore how different cultural tools mediate shared meaning-making and, thus, how discourse is embedded in its specific context (Mercer et al. 2009). Gee and Green's (1998) discourse analysis and Linell's (1998) dialogical approach to communication offer valuable conceptual and analytical tools for researching collaborative activity from the socio-cultural perspective. These approaches target attention on the dynamic and interpretive nature of participants' actions and discourse, and how through these actions and discourse the participants both construct and reflect the context of their activity (Gee and Green 1998). Thus, the methodology presented in this chapter builds on the notion of contexts and situations as being socially constructed (Erickson and Shultz 1981; Linell 1998). According to this view, a context is not a predefined or objective environment (Linell 1998), but only includes those contextual dimensions which are or become relevant to the participants in the activity (Erickson and Shultz 1981; Linell 1998). In this way, the participants themselves create the context through discourse by reflecting and relying on the relevant contextual resources (Linell 1998) or aspects of the situation (Gee and Green 1998) in their joint activity (Goodwin and Duranti 1992; Linell 1998).

Linell's (1998) notion of contextual resources and Gee and Green's (1998) notion of aspects of situation can be seen as possible resources that participants use in their shared meaning-making and interpretation process in discourse and activity. *Immediate and concrete resources or material aspects of situation* refer to the immediate (perceptual) environment which includes, for example, physical spaces, persons, objects and artifacts that are present (potential resources) or referred to (relevant resources) during interaction (Linell 1998). Another immediate resource is Linell's (1998) notion of co-text, which comprises the participants' previous actions and discourse that is actively used in the "new act of sense making" (p. 132). *Mediated and abstract resources or aspects of situation* are reflected and constructed through participants' discursive activity. These include personal, social and cultural knowledge (Gee and Green 1998), such as prior knowledge, experiences, assumptions or beliefs about the things discussed in the discourse in question or about other persons involved in the discourse (Linell 1998). These also include identities which refer to norms and expectations, roles and relationships, and rights and obligations that are relevant in the situation (Gee and Green 1998). Mediated resources also consist of a specific institutional context with its norms, values, regulations and hierarchies as well as an abstract situation definition or the framework of "what is going on" in the actual situation (Gee and Green 1998; Linell 1998). Additionally, Gee's and Green's (1998) semiotic aspect of a situation refers to sign systems such as speech, writing, reading, images and gestures. The context constructed involves these immediate and mediated resources or aspects of the physical, social, cognitive and cultural environments that actualize as relevant through the participants' activity and discourse.

In this chapter I will firstly explore the notion of context through a discourse analysis of collaborative activity in computer-mediated, face-to-face settings. This discursive approach to interaction (Gee and Green 1998) demonstrates how through language and discourse, meaning is carried through time, and how past, present and future contexts are constructed and reflected in the interaction between students as they are engaged in shared activity. Secondly, I will present a multidimensional coding scheme developed for analyzing the contextualized process of collaborative knowledge construction during an asynchronous web-based discussion. This analysis combines the discursive approach with the thematic and functional analysis of discussion and shows how multilevel analysis can be helpful in understanding the reasons behind different collaborative activities. I will also discuss the strengths and weaknesses of the methodology presented.

2.2 Evaluating Collaborative Activity in Its Context

2.2.1 Discursive Approach to Studying Context in Students' Collaborative Activity: The Case of Face-To-Face Activity

In this section, I will draw on data examples reported in the work of Arvaja (2008). The focus of the study was, by means of discourse analysis (Gee and Green 1998), to identify different aspects of the immediate and mediated contexts that were reflected in a student pair's discourse while they were participating in a collaborative task in a higher education context. The subjects of the study were two higher education students studying in a course of educational psychology entitled "Learning Environments and Educational Technology". The pair worked on a project work assignment which concerned the making of an evaluative questionnaire for users (teachers and students) of a web-based learning environment in use at their university. This chapter concentrates on presenting the analyzed extracts from one of the lessons spent on this project work and draws on empirical examples from the classroom discourse. The task in this particular lesson was to revise and continue the development of the questionnaire, which was drafted beforehand, outside the lessons. The students worked in front of a laptop, where they had access to the drafted questionnaire and the web-based learning environment to be evaluated. Transcribed video and audio data were analyzed through an ethnographically grounded approach to discourse analysis (Gee and Green 1998). According to this approach, language is seen as a socio-cultural practice and social resource of groups, and the focus of analysis is more on what participants accomplish through their discourse rather than on what the form or function of the language is, as such. This perspective on language and discourse sees them as simultaneously reflecting and constructing the situation in which they are used. A particular area of interest was in analyzing which aspects of situation (Gee and Green 1998) and contextual resources (Linell 1998) were reflected in the discourse and, thus, what contexts were considered relevant and built into the shared activity.

2.2.1.1 Different Contexts in Meaning-Making

In the lesson analyzed, the particular focus of interest was on the task context of the students' activity and the aspects of situation and resources that were relevant from that perspective. Three different general contexts characterized the data:

- *Immediate and concrete (perceptual) context* consists of any explicit verbal or nonverbal reference to artifact or current discussion that is used in the new act of sense-making (co-text). In this study it also includes semiotic aspects of a situation, such as gesturing, writing and reading.
- *Socio-cultural context* includes references to prior knowledge manifested, for example, in conceptualizations and ideas in planning the questionnaire. Also institutional norms and identities reflected in the situation are part of this context.
- *Local context* includes the immediate 'task frame' that guides students' activity (making a questionnaire) in the moment-to-moment interaction as well as references to past and future activities concerning that activity. Local context creates continuity from lesson to lesson and thus forms the local history (Mercer 2008) of the task.

These contexts and their specific aspects (Gee and Green 1998) and resources (Linell 1998) are now demonstrated through empirical examples and their interpretation through discourse analysis. On the laptop monitor, the drafted questionnaire and the on-line environment to be evaluated were the most important concrete resources that mediated student discourse and activity. The following example (Excerpt 2.1) demonstrates how the drafted questionnaire served as a

Excerpt 2.1 Different contexts in discourse

1. Katrin: Ok. *How often do you use the environment* (reads, whispering, from the screen)?	Immediate context – written text as co-text, reading
Maybe, then, we should start with *this one; which tools are you actually using* (gestures at the screen and reads)	Immediate context – written text as co-text, gesturing, reading
2. Eva: Yhy	–
3. Katrin: And *which ones are you not using* (reads from the screen)?	Immediate context – written text as co-text, reading
4. Eva: *This one* (points to the screen)	Immediate context – written text as co-text, gesturing
5. Katrin: Yeah. I think we should ask may be *both*	Immediate context – previous discussion and written text as co-text
6. Eva: Yeah, yeah. How about *this one? What are the difficulties* (reads from the screen)?	Immediate context – written text as co-text, reading
7. Katrin: Maybe we should leave *an open-ended question* rather to the end	Immediate context – written text and previous discussion as co-text, referring to the question, “What are the difficulties?”, Socio-cultural context – knowledge on questionnaire methodology
8. Eva: But we. How are we going to *rate these ones?*	Immediate context – written text as co-text, Socio-cultural context – knowledge on questionnaire methodology
9. Katrin: *Rate?*	Immediate context – previous discussion as co-text
10. Eva: Yeah	–
11. Katrin: You mean *the last one?*	Immediate context – previous discussion and written text as co-text
12. Eva: No, no *the questionnaire*, are they *open-ended* or…?	Immediate context – written text and discussion as co-text, Socio-cultural context – knowledge on questionnaire methodology
13. Katrin: No, we should make *categories*	Immediate context – discussion as co-text, Socio-cultural context – knowledge on questionnaire methodology
14. Eva: Yhy	–
15. Katrin: Maybe we should. I don’t know	–
16. Eva: *Easy, not easy, stuff like that*	Immediate context – discussion as co-text, Socio-cultural context – rules of categorization
17. Katrin: Yeah. Our *first question: how often do you use it* (reads from the screen)?	Immediate context – written text as co-text, reading

(continued)

Excerpt 2.1 (continued)

18. Eva: *Very often, often, stuff like that*	Immediate context – discussion as co-text, Socio-cultural context – rules of categorization
19. Katrin: Yeah, but maybe we should *specify it. Like each day, twice a week, once a week, once a month or something*	Immediate context – discussion as co-text, Socio-cultural context – rules of categorization
20. Eva: So I put *this* (starts to revise the questionnaire)	Immediate context – written text as co-text, writing
21. Katrin: Maybe we should put *categories* already in brackets, if we know them (follows Eva's writing)	Immediate context – discussion and written text as co-text
Maybe. I don't know; shall we agree *five categories* always?	Immediate context – discussion as co-text, Socio-cultural context – rules of categorization

mediating tool and reference point, enabling the student pair to make progress in the task. Since the nature of the written text is that of a permanent artifact, it gives the students a chance to review and revise the questionnaire through dialogue and other semiotic means and consequently, simultaneously, affects the nature of the dialogue being mediated by the artifact. The example also reflects the socio-cultural context in the form of prior knowledge:
The above excerpt demonstrates how the written text – the drafted questionnaire – simultaneously provides a focus for discourse and reasoning for the task (Turns 1–19), and also embodies the progress made (Turns 20–21). Thus, the questionnaire serves as an 'improvable object' for the students (Wells 1999). It is also used as a shared (concrete) reference object which mediates student discourse. The written text that is explicitly and implicitly referred to in the discourse serves as a co-text for the students, and is actively used in the "new act of sense making" (Linell 1998, p. 132) (Turns 1–8). The discussion itself is also used as a co-text – that is, the previous discussion serves as a co-text for further discussion on the subject at hand. This becomes evident when students start to think about the possible categories and develop their ideas based on one another's suggestions (Turns 9–19). Thus, they are not using the written text as a reference here, but instead are building their ideas on each other's (verbal) suggestions. Co-text (Linell 1998) is a relevant indicator of a shared immediate context being built, and also its prerequisite. In their activity, the students build a shared (immediate) context through different semiotic aspects (Gee and Green 1998). In addition to dialogue, reading and writing, as well as gestures towards the monitor, all are semiotic means that serve in creating the shared context and content-base in the students' collaborative task.

The above excerpt clearly demonstrates the *dynamic nature* of discourse (Mercer 2008). It shows not only how the immediate context is constructed through verbal and nonverbal communication (Linell 1998; Gee and Green 1998), but also how discourse emerges and how speakers' contributions are contingent on what the other speakers say or do (Mercer 2008). The excerpt also embodies a *historical element* (Mercer 2008). Excerpt 2.1 demonstrates how the students draw on prior knowledge

Excerpt 2.2 Different contexts in discourse

1. Katrin: Yeah. I would, I would also put *it* in categories, because *it's easier for us to do the statistics*	Immediate context – written text as co-text, Socio-cultural context – knowledge on doing research
2. Eva: Yhy. Yeah. But how can we do *this why-part?*	Immediate context – referring to the questionnaire in the screen, written text as co-text
3. Katrin: *Why*? I think *we have to categorize it afterwards, like qualitative*	Immediate context – discussion as co-text, Socio-cultural context – knowledge on doing research

on 'questionnaire methodology' in their discourse. Their prior knowledge both implicitly and explicitly stated during the discussion indicates that in evaluative questionnaires one should prefer categories instead of open questions (Excerpt 2.1, Turns 8–13). 'Rules of categorization' are implicitly referred to (Excerpt 2.1, Turns 16, 18, 19) and explicitly stated, such as "Shall we agree five categories always?" (Excerpt 2.1, Turn 21), thus implicitly referring to a Likert-type scale. The students also use their prior procedural knowledge on 'doing research' in their discourse and activity (Excerpt 2.2):

In Excerpt 2.2, Katrin differentiates between quantitative (implicit reference, Excerpt 2.2, Turn 1) and qualitative methods (explicit reference, Excerpt 2.2, Turn 3), referring to her knowledge on research activity. This framework, consisting of rules for making a questionnaire and procedural knowledge about doing research, can also be seen as a previously constructed and learned *cultural model* (Gee and Green 1998) that the students draw on to guide their discourse and activity in the current situation. Cultural models are like theories of action situated in social and cultural experiences (Black 2007), or general ground rules for organizing shared activities (Linehan and McCarthy 2001). Thus, certain implicit rules which are made explicit through discourse guide the students' activity and enable progress in the discussion (see Excerpts 2.1 and 2.2). This is seen, for example, in the progress the students make during their discussion on developing the categories and specifying them (Excerpt 2.1, Turns 13–21). According to Stahl (2004) the building of shared understanding involves making tacit knowledge explicit, which can then become a context for a new object of discussion and its understanding. For example, in Excerpt 2.1, students' prior knowledge of questionnaire methodology is tacit knowledge that is made explicit through the application of the rules it constitutes, and this creates a context for discussion and for developing the object (content) of discussion further.

In addition to prior knowledge, the next excerpt (Excerpt 2.3) shows how also identity, another socio-cultural context, is mediated through student discourse. Also local context in a form of implicit references to past task-related activities is demonstrated:

Excerpt 2.3 Different contexts in discourse

1.	Katrin: Then *he is really interested*	Local context – past discussion with the professor
	that one, that what could be improved	Immediate context – written text as co-text
	[…]	–
5.	Katrin: Maybe, I mean we shouldn't do too many questions, I would say like one…	–
6.	Eva: Yeah	–
7.	Katrin: …one page, maybe we should leave *it* away and just ask *what could be improved*	Immediate context – previous discussion as co-text, referring to question "Do you have difficulties (it) and What could be improved?"
8.	Eva: Oh, okay	–
9.	Katrin: I don't know	–
10.	Eva: Yeah, no problem	–
11.	Katrin: I mean, oh what is more important, we have to use *this one*, 'cause *I think aa, professor is really interested in that one* (points to the screen)	Immediate context – referring to the question "What could be improved?" (this one – that one), written text as co-text, gesturing, Local context – past discussion with the professor, Socio-cultural context – student identity

In Excerpt 2.3, Katrin puts great emphasis on the professor's preferences concerning the questionnaire: "Then *he is really interested*" (Excerpt 2.3, Turn 1) and "What is more important, we have to use this one, 'cause *I think aa, professor is really interested in that one*" (Excerpt 2.3, Turn 11). Thus, when they negotiate what questions to include and what to exclude, Katrin feels that the professor's preferences should override the ones they have ("more important"). This signals certain norms, expectations and obligations in the student-teacher relationship and thus reflects socially valued ways of thinking and acting in the present context (Wells 1999). It implies status differences where the professor has strong authority over the students and their preferences. This is also a good example of how, through discourse, Katrin not only reflects but also re-produces institutional norms and values of the community wherein she acts. Hence, this example also reflects identity applied in the situation (Gee and Green 1998). Katrin's discourse reflects that she identifies with her assumed place in the student-teacher community. According to Wells (2007), identity construction is ongoing and occurs in the situated actions and discourses in which participants engage. In new situations one might apply multiple identities originating from various communities of practice whose values and scripts define our identities. In this case, the student identity is supported by the institutional context wherein the students act. It therefore reflects the previous experiences, attitudes and meanings the students have attached to the activity through their extended participation in this relevant (learning) community (Arvaja 2007a; Crook 1999).

Excerpts 2.2 and 2.3, described above, demonstrate also how the local context provides continuity between different contexts. This means that in the shared task context, the references to past and future discourse or activities concerning the task are relevant in order for one to proceed and understand the task at hand. The local context can be seen as a task frame guiding the activity in the immediate moment-to-moment interaction and influencing the goals and choices made in the immediate context. For example, Katrin suggests creating subcategories for one of the questions because “it’s easier for us to do the statistics” (Excerpt 2.2, Turn 1). Thus, she is referring to future activity concerning the task at hand, which influences the decision-making in the immediate context. In Excerpt 2.3, students are reviewing the questions in the drafted questionnaire and negotiating what should be excluded. Katrin suggests excluding the last question, “Do you have difficulties?”, and suggests including the question “What could be improved?” (Turn 7). She justifies this suggestion by referring to a past discussion she had with the professor teaching the course (Turns 1, 11). In the analyzed lesson the references to the past and future activities are typical justifications for decision-making and negotiation in the moment-to-moment interaction. This demonstrates how the participants’ task frame crosses time and events outside the immediate perceptual context (Mercer 2008).

From the point of view of collaboration, the focus of the presented analysis was both on the historical as well as the dynamic aspect of discourse (Mercer 2008). The historical dimension of the discourse was apparent in that the interaction was located within a particular institutional and cultural context. In the analyzed examples, it was demonstrated how students drew on some past experience or prior knowledge that was used as a resource for building understanding in the present situation and how they applied ‘a cultural model’ (Gee and Green 1998) of doing research in their activity. Discourse also reflected certain norms, expectations and obligations that were socially valued in the present context (Wells 1999). The dynamic aspect of discourse became evident from the way in which the written text, previous discussion or speech turn served as a co-text in student meaning-making, as a ground for building on one another’s suggestions, and reasoning further the subject at hand. The notion of intertextuality provides another conceptualization for understanding how different moments in time were tied together and how the students drew on past texts to construct present texts and implicate future ones (Gee and Green 1998; Pappas et al. 2002; Staarman et al. 2004). Furthermore, Grossen’s (2009) notion of “dialogicality”, with its spatial and temporal dimensions, provides another tool for understanding how every situation has a “here-and-now” (spatial) and a “there-and-then” (temporal) dimension. Often the analysis of collaborative interaction, e.g., in the socio-cognitive research tradition, is interested in analyzing the “here-and-now” situation, leaving out the historical and temporal nature of discourse (Grossen 2009; Mercer 2008).

Even though the analysis presented above was able to demonstrate the embeddedness of different contexts in students’ activity, the analysis was not, as a whole, temporally (Mercer 2008) or chronologically (Hmelo-Silver et al. 2008) ordered. Thus, even though the analysis showed how through discourse different historical timeframes or layers (past, present, future) were present in the immediate context of activity through intertextual referencing, the analysis did not focus on analyzing the temporal history of the whole data. What the students actually learned through participating in extended dialogue with each other and the teacher during the course cannot be discussed based on the short data excerpts presented here. Temporal analysis of the whole discourse would have helped to identify how students' ideas developed and changed through the extended process of interaction in the group, and how possibly new concepts, ways of thinking and solving problems were appropriated (Mercer 2008).

2.2.2 Contextual Process of Collaborative Knowledge Construction: The Case of Asynchronous Web-Based Discussion

Even though the analysis presented above clearly demonstrated how students' collaborative activity was embedded in the immediate and mediated contexts where the collaboration took place, and therefore warned against analyzing (collaborative) activity out of its context, it provided few tools for evaluating the *collaborative* aspect of interaction and knowledge construction. Next, I will illustrate a methodology developed for studying students' collaborative knowledge construction activity in an asynchronous discussion forum, not forgetting the contextual nature of interaction and shared meaning-making. The analysis and results reported in Arvaja (2007b) will be used as a basis for demonstrating and discussing the methodology.

The subjects of the study consisted of two small groups of teacher education students studying the pedagogy of pre-school and primary education in a web-based learning environment. The students were set a so-called 'open problem', meaning that they had to create and solve a problem relating to the theme "Differentiation in teaching reading". The students' task was to discuss in an asynchronous discussion forum the problem they had created and finally to prepare a lesson plan for teaching reading. The two groups were chosen for detailed analysis and comparison from among seven groups engaged in the same basic task, because they had created the same problem: "How to differentiate teaching reading in a classroom where pupils are on different levels as regards reading ability".

The study concentrated on examining the asynchronous web-based discussion that each of the groups had. One focus of the framework developed for the analysis was on the functional analysis of communication, which shed light on the purpose of the discussion. The communicative functions were adapted from the framework for analyzing language functions developed by Kumpulainen and Mutanen (1999). However, these language functions were not used as predefined categories; rather, the specific context of the data was taken into account in interpreting the function of communication. Thus, the communicative functions were contextual in nature, depending on the topic of the discussion and the interpretations made by the participants involved in these discussions. The functional analysis of the web-based messages focused on the purposes for which language was used in the given context. The communicative functions were identified by their content and form and by their effect on and relation to the discourse of which they were part. From the data, 11 functions of communication were found*; interrogative, responsive, judgmental,* *evaluative, suggestive, informative, exemplary, elaborative, justificational, reasoning* and *summarizing* functions (Arvaja 2007b).

However, the functional analysis, as such, does not tell much about the content and nature of knowledge construction (Crook 1999; Stahl 2002). Thus, in order to truly evaluate the *shared knowledge construction,* we actually have to examine both the kind of knowledge that is constructed as well as whether the knowledge is mutually constructed. Therefore, the method developed focused also on the matic content of the discussion as well as on the contextual resources (Linell 1998) used for knowledge construction. *Content analysis* of the messages was conducted to explore the thematic network of the messages: What knowledge or information was dealt with in each message? How were the messages thematically related to each other? Thus, the unit of analysis was the thematic meaning unit rather than the message, paragraph or sentence as such (e.g., Rourke et al. 2001). After reading the messages through a number of times, two broader themes were identified in the discussions of both the groups: "Methods for teaching reading" and "Differentiating activities in teaching reading". Thematic content analysis was also conducted at the utterance/several utterances level to identify the main (sub) themes of discussion within the two broader discussion themes. The notion of contextual resources (Linell 1998, see Sect. 2.1) was used as an analytical tool in studying what aspects of the potential context the students made relevant in the process of shared meaning-making. Relevant resources were those referred or oriented to in the discourse. From the data, five broader categories of resources were found (Arvaja 2007b):

- *Course material*: In discourse, students refer directly, for example, to lectures or articles which serve as theoretical background material for the task, or the discussion may be identified as being based on the course material.
- *Own idea*: In discourse, students use their own ideas, which are mostly manifested in *action* and *activity descriptions*. Own ideas are usually based on common knowledge about school practices.
- *Own conception*: In discourse, students use their own conceptions of either practical or more abstract issues or knowledge manifested in *interpretations* of issues or knowledge (e.g., the consequences of a practical suggestion or the conceptualization of theoretical knowledge). In discourse this shows in reasoning and justifying.
- *Own experience*: In discourse, students refer to their own teaching experiences manifested mainly in *case descriptions* or *examples.* References to own teaching practices also reflect the *identity* (Gee and Green 1998) applied in the situation.
- *Co-text*: Co-text refers to the fact that students build their thoughts on other students' thoughts. In discourse, students directly or indirectly refer to *interpretations, case and activity descriptions or examples presented by others* by developing them further.

Thus, to analyze collaborative knowledge construction, a multidimensional coding scheme was developed. The analysis of the communicative functions and contextual resources was limited to students' content-based activity. In another study, the whole data was coded and new categories of communicative functions and contextual resources were found, such as the social and organizational functions and a document base as a resource to refer to (see Arvaja and Hämäläinen 2008). However, this new analysis did not affect the categories developed for content-based activity presented in this paper. In addition to the analysis of the functions, resources and themes of discussion, for the purposes of this paper, the discursive features (Gee and Green 1998) of the discussions of the two groups were analyzed to deepen the interpretations made based on the coding category.

2.2.2.1 Analyzing Collaborative Knowledge Construction in Its Specific Context

To exemplify the multidimensional coding scheme and to give a contextualized interpretation of the students' activity (Hicks 1996), a detailed analysis is demonstrated in the next table (Table 2.1) based on the discussion of Group A.

In first three of the messages (Messages 7, 8 and 11), the students are wondering how they should differentiate the teaching to fit pupils' abilities (theme: Differentiating activities in teaching reading). From the thematic content and functions of communication we can see that Iina first suggests dividing the pupils into ability groups, but also reasons that they ought to have some joint lessons as well (Message 7). Otto elaborates this further by suggesting peer teaching in mixed groups (Message 8). In the messages the students also negotiate how the pupils should be taught in "ability" groups. Iina suggests using a school assistant (Message 7) and Alisa agrees (Message 11). Alisa elaborates further on differentiation by suggesting the use of a special teacher for dyslexic and remedial instruction for weaker pupils (Message 11). In the three messages the students mostly elaborate one another's ideas and justify their own and others' suggestions. They also ask for confirmation for their own ideas (Interrogative; Messages 7, 11). In messages 20 and 21, the students are discussing the method that would be best for pupils who have difficulties. This becomes a real problem-solving situation for them. Alisa asks for clarification on the KÄTS method while at the same time elaborating on its features (Message 20). Iina replies by confirming and justifying the selection by describing the features supporting the selection of that particular method (Message 21).

From the contextual resources we can see that in the 'differentiation' theme the students build their discussion mostly on one another's (co-text) and their own ideas and conceptions. In the 'methods' theme the students draw also on course material. The notion of co-text is particularly important from the point of view of collaboration, because it shows whether the students build their thoughts and discussion on other students' thoughts and discussion (Arvaja 2007b). Thus, it shows whether the theme/s under discussion is co-constructed. This means that in discourse, students directly or indirectly refer to descriptions, interpretations, or examples presented by others by developing them further. It is important to note that although the elaborative function of communication indicates that students develop previously offered knowledge further, the concept of co-text additionally indicates whether some other functions, such as asking and answering questions, justifications and reasoning, are built on one

Table 2.1 Empirical example of the analysis of Group A discussion by using a multi-dimensional coding scheme

Message (number, title, sender)	Function of communication	Contextual resources	Main thematic content	Discursive features
(7. More about the task…, Iina)				
Would it be good idea to divide the pupils into "ability groups" and for every group think of what method to start with?	Suggestive, interrogative	Own idea	Differentiating by the means of grouping on ability	"Ability groups"
Could we have, for example, a school assistant in our classroom (could be) so differentiating the pupils would be easier?	Suggestive, justificational, interrogative	Own idea	Differentiating by the means of assistant	–
We still need to think how often we should have these kinds of "ability groups", I suppose not every lesson, I suppose we should have some joint teaching as well…	Reasoning	Own conception	Mixing different ability students	"Ability groups" Hesitation common sense
(8. Re: More about the task, Otto)				
As Iina said, it isn't necessary to differentiate the pupils in every lesson	Judgmental	–	–	–
It is probably good for the dyslexic as well to be with "normal" learners	Justificational	Own conception Co-text	Mixing different ability students	Hesitation Common sense "Normal" learners
Those who can already read could read in small groups and those who are still trying to acquire the ability to read would listen and would see the others act	Elaborative	Own idea Co-text	Peer teaching in mixed groups	–
This way all could participate on their own level. I believe it also motivates the more skilful pupils. Otherwise they escape the teacher's attention	Justificational	Own conception	–	Skilful pupils Common sense
(11. Thoughts, Alisa)				
–School assistant available	Summarizing	–	–	Ability groups
–Ability groups in teaching every now and then. […]				
Dyslexic children and those who have difficulties in perceiving text and letters will attend a special teacher, I guess	Elaborative	Co-text Own idea	Differentiating by the means of special teacher	Hesitation Common sense

(continued)

Table 2.1 (continued)

Message (number, title, sender)	Function of communication	Contextual resources	Main thematic content	Discursive features
It also occurred to my mind that if a teacher gives remedial instruction to a weaker pupil, one could in advance go through the section etc. to be handled together in the next lesson, already in the remedial session	Elaborative	Co-text Own idea	Differentiating by the means of remedial instruction	Weaker pupil
In this way the weaker pupils could experience feelings of success, as they don't always have to be the slowest ones	Justificational	Own conception	–	Weaker pupils Common sense
(20. Hello, Alisa)				
It's very difficult to decide what to do with those who have difficulties	Judgmental	–	Method for those who have difficulties	–
It is anyway the case that these children have difficulties in many different areas…	Reasoning	Co-text Own conception	–	Common sense
I cannot offer any method; the only thing that comes into my mind is that the teacher supports every child and tries to help them forward	Responsive	Co-text Own idea	–	Hesitation (uncertainty) Common sense
Was it the case that the KÄTS method was initially developed for children who have difficulties in reading, or is my memory at fault?	Interrogative Elaborative	Course material Co-text	Specific features of KÄTS method	Theoretical ground Hesitation (uncertainty)
If it is so, could we apply KÄTS in this group as well?	Suggestive Interrogative	Own idea	–	–
(21. Re: Hello, Iina)				
Yeah, it was the case that KÄTS was initially for those who had difficulties in reading…if I remember correctly	Responsive	Co-text Course material	Specific features of KÄTS method	Hesitation (uncertainty)
Let's start with that method, I can't figure out anything else	Judgmental	–	–	Hesitation (uncertainty)
And KÄTS is a method that deals with these letters and speech sounds and their perception, which I think is important for those with dyslexia	Justificational	Course material Co-text Own conception	Specific features of KÄTS method	Theoretical ground

another's ideas or thoughts, thereby developing the groups' reasoning further, or whether it is more a case of the students sharing knowledge (no co-text) and following their individual paths of reasoning (e.g., Barron 2000). Communicative functions, such as judgmental or evaluative comments, are not considered as co-text, even though the previous discussion is referred to. This is because the previous suggestion, for example, is only acknowledged, but not developed further (Arvaja et al. 2007).

As the activity in this study took place in the virtual environment, the line between immediate and mediated context is more difficult to draw than in the previous case. Roughly, one could say that only co-text is part of the immediate context; other resources are more or less mediated. The only semiotic means (Gee and Green 1998) in the web-based discussion are reading (others' messages) and writing (own messages). Thus, the written text (others' messages) that represents (text-based) discussion serves as a co-text for the students (see Sect. 2.2.1.1 for comparison). Own ideas, conceptions and experiences are usually related to prior knowledge the students have on reading methods and differentiation, and thus represent more the socio-cultural context of an activity. Also conceptualizations made based on the previously read texts or listened lectures (course material) can be considered as mediated resources. However, the students might use the course material also as a concrete resource while working on their computer. Thus, assigning resources into different categories or into immediate or mediated resources is, in some cases, more technical than actual, as resources "overlap considerably" (Linell 1998, p. 132). As is demonstrated in another study based on this data (Arvaja and Hämäläinen 2008), the local context is manifested in the students 'organizational talk' that ties together the problem-solving discussion and the document writing activity (lesson plan). Task frames (interpretations and goals of the task) "selecting the method" and "means to differentiate" provide another local context that guides the discursive activity throughout the discourse (see Arvaja 2007b).

The discursive features of discussion are examined to interpret deeper meanings and reasons for the students' activity. From the discursive features we can see that the task is challenging for the students and they are uncertain and hesitant in their suggestions. Some discursive features of their communication indicate their hesitation: "I suppose" (Message 7), "It is probably", "I believe" (Message 8), "I guess" (Message 11). Also in messages 20 and 21, students are directly referring to their uncertainty: "or is my memory at fault?", "if I remember correctly", "I can't figure out anything else". The high frequency and also the nature of questions – mostly in the form of requests for confirmation of one's own suggestion – are further evidence of this uncertainty. Under the methods theme, the fact that the students are justifying or reasoning their conceptions theoretically indicates that the methods for teaching reading are something of which the students in this group have no experience. Thus they have to lean on theoretical course material and one another (co-text) in their reasoning. As they do not have experience in applying the methods, the methods become something they try to remember ("if I remember correctly"), not necessarily something they understand. Their discourse resembles a kind of 'in theory' talk. As for the differentiation theme (Messages 7, 8 and 11), in turn, the students are mostly using 'common sense' in their justifications and reasoning the issue: "I believe it also motivates the more skilful pupils". It is also present in their discussion on the method theme: "I cannot offer any method; the only thing that comes into my mind is that a teacher supports every child and tries to help them forward" (Message 20, Table 2.1). This discourse could be described as a kind of 'common sense' talk in the sense that it helps the students to make inferences about what the other one means, and they rely on the expectation that it makes sense, and thereby also helps them avoid conflicts and disagreements.

2.2.2.2 Combining Quantitative and Qualitative Analysis

In the next table (Table 2.2), the percentages for the five main functions of communication and contextual resources are presented for both of the themes to demonstrate the overall nature of collaborative activity in both of the groups.

Quantifying the categories of communicative functions and contextual resources serves as a valuable tool for comparing the general differences and similarities between the two groups. However, linking the analysis with qualitative description and interpretation of the messages, which also takes into account the thematic content and discursive features of discussion, enables a deeper understanding of the ways in which the dimensions interact in the situation. From the contextual resources we can see that for the theme 'methods', the main resources of Group A were co-text and course material (Table 2.2). As the analysis (Tables 2.1 and 2.2) demonstrates, as far as the methods were concerned, the students faced a real problem-solving task, and needed the course material (31%) and one another (co-text, 37%) in order to solve the problem of "what methods to select". A thematic analysis showed that the students discussed all the three methods thoroughly (see Arvaja 2007b):

> *We could use LPP method at least for shared teaching, because the pupils don't need to be able to read at all. This would practise that visual perception and the perception of wholeness, which is important for everyone, no matter what phase of reading they are.*
>
> (Message 12, Iina, Group A)

As this example again shows, the students relied on theoretical justifications, 'in theory' talk, in their discussion. However, the reliance on theory as 'something to remember', as was pointed out earlier, is again confirmed, because the student here is actually talking about the features of the KPL method, not those of the LPP, thus making a misinterpretation. As can be seen from the contextual resources drawn on in Group B, the students relied mostly on their own experience (42%) in choosing the method (Table 2.2). The next example illustrates the nature of their discussion:

> *I myself have used the KÄTS method, but I have noticed that if you have dyslexia, you have to slide letters close together and then it almost changed the old a-u au t-o to auto [car in Finnish]...*
>
> (Message 8, Jaana, Group B)

Whereas Group A discussed all of the methods thoroughly before making the selection, Group B did not discuss any other methods, but chose the KÄTS

Table 2.2 Five main functions of communication and contextual resources in Group A and B discussions (adapted from Arvaja 2007b)

	Methods theme		Differentiation theme	
	Group A (%)	Group B (%)	Group A (%)	Group B (%)
Communicative functions				
Elaborative	16	–	23	14
Exemplary	–	11	–	12
Informative	12	–	–	–
Interrogative	16	17	23	–
Judgmental	–	–	13	14
Justificational	16	22	13	18
Suggestive	12	34	13	17
In total	72	84	85	75
Contextual resources				
Co-text	37	0	38	24
Course material	31	8	3	2
Own conception	13	33	21	22
Own experience	0	42	0	32
Own idea	19	17	38	20
In total	100	100	100	100

method that they all knew, based on their teaching experience. The students used their own experiences in legitimizing the selection of the method instead of a critical comparison of the different methods: “I don’t properly know any other method than KÄTS, but I don’t mind what methods we select for this class…” (Message 17, Jaana, Group B). Thus, whereas Group A students justified the selection of the methods on an ‘in theory’ basis, Group B students based their justifications on an ‘in practice’ basis. The main function of communication in Group B was suggesting (34%), and, as the total lack of co-text (0%) demonstrates, the suggestions made were not related to other students’ suggestions. Thus, they did not regard one another’s contributions as relevant resources in solving the problem at hand.

As regards the ‘differentiation’ theme, the groups were more similar in quantitative comparison (Table 2.2). Of the five main functions of communication, four were the same. Co-text was also among the main contextual resources drawn on in both of the groups. This shows that the knowledge was co-constructed in both of the groups. Lack of theoretical knowledge (3/2%, Arvaja 2007b) and references to one’s own ideas (Group A, 38%) and experiences (Group B, 32%) indicates that the knowledge construction was based on common practical knowledge. In fact, the thematic analysis showed that both of the groups handled exactly the same sub-themes as a means to differentiate teaching (see Arvaja 2007b). This is an indication that all of the students shared cultural knowledge (Gee and Green 1998; Linell 1998) about these school practices, even though the ‘level of this knowledge’ differed, as can be seen by examining the discursive features. As was demonstrated above, Group A students used ‘common sense’ talk in justifying the suggestions made as regards the means to differentiate the teaching. However, Group B students' discussion had more features of 'realistic' talk.

> *You can also differentiate pupils according to remedial instruction, but I have noticed that with the existing time frames it is impossible as the only way, as you can give remedial teaching for an hour or two in months.*
>
> (Message 6, Mari, Group B)

Both this (Message 6) and the previous example from the method theme (Message 8) show that Group B saw some problems in real-life activity relating to the practices discussed, and they critically evaluated the established practices based on their experience and the reality they had faced in the field. Thus, both groups shared the same common knowledge, but Group B had a deeper understanding of the issues discussed, based on their experience, and this showed in 'realistic' talk as opposed to 'common sense' talk in Group A. Another difference in discursive features was that whereas Group B students used professional terms in their discussion, Group A students used more novice-like terms. Table 2.3 summarizes the differences in discursive features between the groups.

Combining the three dimensions of the analysis – the theme and function of communication and the source of knowledge – enabled an evaluation of what kind of knowledge was constructed, how it was constructed, what resources were used for it and whether the knowledge was co-constructed. Even though the analysis of communicative functions and contextual resources was performed at the utterance level and hence represented quite a static kind of analysis (Grossen 2009), it was still able to some extent to capture the dynamic nature of interaction (Mercer 2008). Most of the communicative functions, such as elaboration, answering, summarizing, evaluating and judgment, imply the connection to the content and function of the previous message(/s). In addition, the contextual resource co-text indicated whether the content of the previous message was used as grounds for thinking and developing the knowledge further (co-construction), and not just for acknowledging the previous thoughts presented. Quantifying the analysis of communicative functions and resources offered valuable direct knowledge on the general similarities and differences between different groups. Contextual resources indicated whether the knowledge discussed was based on practical or theoretical knowledge, that is, on one's own experiences as a teacher, ideas based on common knowledge, or on course material. Communicative functions, in turn, allowed interpretations to be made of the quality and purpose of discussions in the two groups. Thus, the quantitative analysis

Table 2.3 Discursive features of discussion in Group A and B

Group A	Group B
Hesitation/uncertainty	Confidence
'In theory' talk	'In practice' talk
'Common sense' talk (unproblematizing)	'Realistic' talk (problematizing)
Novice terms: e.g., "ability groups", weaker, "normal", skilful and advanced pupils	Professional terms: e.g., reading groups, challenging pupils and fluent readers

was able to give general knowledge on the nature and ‘level’ of collaborative knowledge construction in the two groups.

However, only a detailed qualitative analysis of the relations among specific thematic content, communicative functions and contextual resources as well as the discursive features of discussion made it possible to gain a deeper understanding of the reasons behind these similarities and differences. This interpretative analysis was able to open up the “dialogicality” of the situation (Grossen 2009): how the meaning of the here-and-now situation (e.g., interpretation of the present task) was understood with respect to there-and-then situations, such as other activities that the students had been involved in (e.g., teaching experiences) or had some representation of (e.g., theoretical conceptualization based on course material), and in what kind of activity and knowledge formation this resulted. Interpretation of all the aspects taken into account in the analysis as intertwined showed, for example, how in Group B the selection of the reading method (the problem at hand) was suggested and justified based on one’s own and shared teaching experience manifested in ‘realistic’ and ‘in practice’ talk, and regarded as sufficient justification. Thus, Group B students were able to pass the task by relying on their former experiences as teachers, rather than relying on one another’s contributions (no co-text) or theoretical material in completing the task. They regarded their own teaching experience as relevant and sufficient (context) for accomplishing the task. Furthermore, the interpretative analysis showed how in Group A an emergent understanding of reading methods, manifesting itself in ‘common sense’ and ‘in theory’ talk, was expressed as hesitation in the situation and led the students to seek confirmation (interrogative) from other students, and consequently led to shared problem-solving (e.g., responsive, co-text) in the situation (see Table 2.1). Thus, being novices, they seemingly needed one another and theoretical course material in completing the task. The qualitative analysis and interpretation of the here-and-now and there-and-then situation illuminated how the students interpreted the task at hand and what resources were realized as relevant, and why (/not). Thus, the analysis was able to provide contextualized interpretations of why certain activities occurred.

2.3 Discussion

One focus of this book is to present different methodologies for studying interaction in various CSCL contexts. If I reflect on the methodology presented in this particular chapter, what was actually sought through the analyses was dialogicality (Grossen 2009; Linell 1998) in situation, rather than interaction in situation. Firstly, in both of the empirical cases, building the shared immediate here-and-now situation or the shared ‘content and activity space’ through co-textual referencing was shown to be a prerequisite for collaborative activity. In such a situation, the students’ activity was coordinated (Barron 2000) and the activity was organized around joint problem-solving efforts manifested in co-construction of solutions and referring to and expanding one another’s ideas. Secondly, the studies also demonstrated how collaborative activity was rooted in the mediated there-and-then situation; in the socio-cultural context and in the history of the students and their activities.

One weakness of the presented methodology is that it, as such, captured more the 'visible' analyst's perspective. Additional data in the form of interviews or diaries, for example, would have shed more light on (other) resources relevant from the students' perspective (e.g., Arvaja 2007a). Another limitation of the methodology as presented here relates to its use of the group as the unit of analysis. However, the study of Arvaja and Hämäläinen (2008) used an individual level of analysis based on the multidimensional coding scheme and demonstrated how the individual students in the group adopted different functional roles and resources in different tasks (e.g., knowledge provider, knowledge elaborator, social supporter). Also, another study (see Arvaja et al. 2007) combined the individual self-report data (questionnaire) with the group-level discourse data to give knowledge on the subjective meanings the students attached to the group's learning activities. The self-report data was able to validate the findings of the discourse data. It is thus obvious that the data from multiple sources as well as individual and group-level perspectives are needed to fully understand the contextual nature of collaborative learning.

Empirical examples illuminated how it is pedagogically important to see that original task designs are always re-interpreted by the learners themselves and the actual realisations may radically differ from what might have been planned by the task designers. For example, in the second case, whereas the task was a real problem-solving task for Group A, the students in Group B faced no challenges as such and merely reproduced their commonly shared knowledge. The task and its structuring in the web-based learning environment were apparently unresponsive to the different resource needs of the groups. The notion of contextual resources (Linell 1998) can thus be used as a tool in evaluating the success of designed collaboration activities by revealing, on the one hand, how students interpret and make use of available resources, and on the other hand, how the resources used (e.g., conceptual, social, material, technological) support students' collaboration. This, in turn, helps to identify critical points for designing learning and teaching activities in flexible ways, both before and during the ongoing activity. From the teacher, this requires appropriate scaffolding (Rasku-Puttonen et al. 2003) and flexible structuring of students' activities, for example, by offering new resources that become relevant for students' activity (Dillenbourg and Tchounikine 2007).

References

Arvaja, M. (2007a). Contextual resources in meaning negotiations of a student pair in a web-based history project. *International Journal of Educational Research, 46*(3/4), 215–228.

Arvaja, M. (2007b). Contextual perspective in analysing collaborative knowledge construction of two small groups in web-based discussion. *International Journal of Computer-Supported Collaborative Learning, 2*(2/3), 133–158.

Arvaja, M. (2008). Discursive approach for studying contexts in students' collaborative activity. In G. Kanselaar, V. Jonker, P. A. Kirschner & F. J. Prins (Eds.), *Proceedings of the eight* *international conference for the learning sciences* (Vol. 1, pp. 52–59). International Society of the Learning Sciences.

Arvaja, M., & Hämäläinen, R. (2008). Collaborative knowledge construction during structured tasks in an online course at higher education context. In G. Ollington (Ed.), *Teachers and teaching: Strategies, innovations and problem solving*. New York: Nova Science Publishers.

Arvaja, M., Salovaara, H., Häkkinen, P., & Järvelä, S. (2007). Combining individual and group-level perspectives for studying collaborative knowledge construction in context. *Learning and Instruction, 17*(4), 448–459.

Baker, M., Andriessen, J., Lund, K., van Amelsvoort, M., & Quignard, M. (2007). Rainbow: A framework for analysing computer-mediated pedagogical debates. *International Journal of Computer-Supported Collaborative Learning, 2*(2/3), 315–357.

Barron, B. (2000). Achieving coordination in collaborative problem-solving groups. *Journal of the Learning Sciences, 9*(4), 403–436.

Black, L. (2007). Analysing cultural models in socio-cultural discourse analysis. *International Journal of Educational Research, 46*(1–2), 20–30.

Bliss, J., & Säljö, R. (1999). The human-technological dialectic. In J. Bliss, R. Säljö, & P. Light (Eds.), *Learning sites: Social and technological resources of learning* (pp. 1–16). Amsterdam: Pergamon.

Crook, C. (1999). Computers in the community of classrooms. In K. Littleton & P. Light (Eds.), *Learning with computers. Analysing productive interaction* (pp. 102–117). London: Routledge.

Dillenbourg, P., & Tchounikine, P. (2007). Flexibility in macro-scripts for computer-supported collaborative learning. *Journal of Computer Assisted Learning, 23*(1), 1–13.

Erickson, F., & Shultz, J. (1981). When is a context? Some issues and methods in the analysis of social competence. In J. Green & C. Wallat (Eds.), *Ethnography and language in educational settings* (pp. 147–160). Norwood, NJ: Ablex.

Fischer, F., Bruhn, C., Gräsel, C., & Mandl, H. (2002). Fostering collaborative knowledge construction with visualization tools. *Learning and Instruction, 12*(2), 213–232.

Gee, J., & Green, J. (1998). Discourse analysis, learning and social practice: A methodological study. *Review of Research in Education, 23*, 119–169.

Goodwin, C., & Duranti, A. (1992). Rethinking context: An introduction. In A. Duranti & C. Goodwin (Eds.), *Rethinking context: Language as interactive phenomenon* (pp. 1–42). Cambridge: Cambridge University Press.

Grossen, M. (2009). Social interaction, discourse and learning. Methodological challenges of an emergent transdisciplinary field. In K. Kumpulainen, C. Hmelo-Silver & M. César (Eds.), *Investigating classroom interaction*. Sense Publishers.

Hicks, D. (1996). Contextual inquiries: A discourse-oriented study of classroom activity. In D. Hicks (Ed.), *Discourse, learning and schooling* (pp. 104–141). Cambridge: Cambridge University Press.

Hmelo-Silver, C. (2003). Analyzing collaborative knowledge construction: Multiple methods for integrated understanding. *Computers & Education, 41*(4), 397–420.

Hmelo-Silver, C., Chernobilsky, E., & Jordan, R. (2008). Understanding collaborative learning processes in new learning environments. *Instructional Science, 36*(5–6), 409–430.

Kumpulainen, K., & Mutanen, M. (1999). The situated dynamics of peer group interaction: An introduction to an analytic framework. *Learning and Instruction, 9*(5), 449–473.

Linehan, C., & McCarthy, J. (2001). Reviewing the "community of practice" metaphor: An analysis of control relations in a primary classroom. *Mind, Culture and Activity, 8*(2), 129–147.

Linell, P. (1998). *Approaching dialogue. Talk, interaction and contexts in dialogical perspectives*. Amsterdam: John Benjamins.

Mercer, N. (2008). The seeds of time: Why classroom dialogue needs a temporal analysis. *The Journal of the Learning Sciences, 17*(1), 33–59.

Mercer, N., Littleton, K., & Wegerif, R. (2009). Methods for studying the processes of interaction and collaborative activity in computer-based educational settings. In K. Kumpulainen, C. Hmelo-Silver, & M. César (Eds.), *Investigating classroom interaction*. Sense Publishers.

Pappas, C., Varelas, M., Barry, A., & Rife, A. (2002). Dialogic inquiry around information texts: The role of intertextuality in constructing scientific understanding in urban primary classrooms. *Linguistics and Education, 13*(4), 435–482.

Rasku-Puttonen, H., Eteläpelto, A., Arvaja, M., & Häkkinen, P. (2003). Is successful scaffolding an illusion? - Shifting patterns of responsibility and control in teacher-student interaction during a long term learning project. *Instructional Science, 31*(6), 377–393.

Rourke, L., Anderson, T., Garrison, D. R., & Archer, W. (2001). Methodological issues in the content analysis of computer conference scripts. *International Journal of Artificial Intelligence in Education, 12*, 8–22.

Staarman, J. K., Aarnoutse, C., & Verhoeven, L. (2004). Connecting discourses: Intertextuality in a primary school CSCL practice. *International Journal of Educational Research, 39*, 807–816.

Stahl, G. (2002). Rediscovering CSCL. In T. Koschmann, R. Hall, & N. Miyake (Eds.), *CSCL 2: Carrying forward the conversation* (pp. 169–181). Hillsdale, NJ: Lawrence Erlbaum Associates.

Stahl, G. (2004). Building collaborative knowing. Elements of a social theory of CSCL. In P. Dillenbourg (Series Ed.) & J. W. Strijbos, P. A. Kirschner, & R. L. Martens (Eds.), *Computer-supported collaborative learning*, *What we know about CSCL… and implementing it in higher education* (Vol. 3, pp. 53–85). Boston, MA: Kluwer Academic Publishers.

van Boxtel, C., van der Linden, J., & Kanselaar, G. (2000). Collaborative learning tasks and the elaboration of conceptual knowledge. *Learning and Instruction, 10*(4), 311–330.

Vygotsky, L. (1978). *Mind and society*. Cambridge, MA: Harvard University Press.

Wells, G. (1999). *Dialogic inquiry. Toward a sociocultural practice and theory of education*. New York: Cambridge University Press.

Wells, G. (2007). Who we become depends on the company we keep and on what we do and say together. *International Journal of Educational Research, 46*(1–2), 100–103.

Wertsch, J. (1991). A sociocultural approach to socially shared cognition. In L. Resnick, J. Levine, & S. Teasley (Eds.), *Perspectives on socially shared cognition* (pp. 85–100). Washington, DC: American Psychological Association.

Chapter 3
Understanding Learners' Knowledge Building Trajectory Through Visualizations of Multiple Automated Analyses

Nancy Law, Johnny Yuen, Wing O.W. Wong, and Jing Leng

Abstract In recent years, knowledge building as a pedagogical approach has gained in popularity, and some sizeable networks of teachers using the Knowledge Forum® platform to support the learning and teaching process have emerged. This provides a unique opportunity for empirical explorations that span across different classrooms to identify trends and regularities in the knowledge building developmental trajectory of a community of learners (including the teacher who is designing and facilitating the educational process) in classroom contexts. On the other hand, it poses a serious challenge to teachers and researchers on how to analyze and make sense of the large corpora of discourses generated by students. This study is an attempt to develop a methodology that can take advantage of the increasing collection of online corpora in Knowledge Forum® for the incremental establishment of profiles and models of knowledge building behavior and interactions that reflect different levels of productivity in knowledge building. The focus of the present research is on the progress and development of the entire community, which aligns with the predominant focus on the construction and advancement of collective knowledge in knowledge building research.

The goal of this study is to develop a methodology, including the identification of appropriate indicators and tools that can be used to provide a quick, first level assessment of the level of knowledge building reflected in a Knowledge Forum® corpus using machine analysis and visualization, building on previous research in the area. Such a methodology will help to (1) build an empirically grounded understanding of learners' trajectories of advancement in knowledge building; and (2) identify pedagogical and facilitation designs that are more conducive to deeper levels of knowledge building by students. We implemented automatic encoding and visualization on a number of Knowledge Forum corpora collected from primary and secondary classrooms in Hong Kong around the themes of energy crisis and global warming.

The findings indicate that visualization of different indicators applied in sequence provides some useful insight on the quality and level of engagement in

N. Law (✉)
University of Hong Kong
e-mail: nlaw@hku.hk

S. Puntambekar et al. (eds.), *Analyzing Interactions in CSCL*, Computer-Supported Collaborative Learning Series, DOI 10.1007/978-1-4419-7710-6_3,

online asynchronous discussions. Basic participation statistics at discourse and thread levels are useful in discriminating truncated inquiries from other, more engaging ones. Linear visualization of the time sequence of discourse markers for argumentation, questions and scaffolds sheds light on whether there is evidence of cycles of progressive inquiry. Fine grained identification of subject matter content reveals whether the discussion is likely to be an extensive inquiry or an intensive one. The study also demonstrates a clear need for automated coding and visualization of discourse corpora to be augmented and validated by a review of the actual discourse data by human researchers to yield more in-depth understanding of the knowledge building processes.

3.1 Introduction

According to Bereiter (2002) and Scardamalia (2002), knowledge building is the process through which knowledge advances in human societies, and that learning can take place in the process. Knowledge building is one of the earliest endeavors that shape the field of CSCL (Koschmann 2003). However, while some research groups in the CSCL community may have a general focus on supporting socially mediated learning, the primary concern in knowledge building is not on individual learning, but the collective advancement of community knowledge and the improvement of ideas (Scardamalia 2002). Scardamalia and Bereiter (2003) further propose that technology plays an important supporting role in the process and that the design features of the technology matter. Knowledge Forum® has been developed as a platform with the key technological features identified as crucial to supporting the socio-metacognitive dynamics necessary for knowledge building.

The knowledge building approach to education focuses on the construction and advancement of collective knowledge (Lamon et al. 2001). Obviously, it is not the expectation that individuals will achieve the same understanding at the end of a collaborative process. However, just as the productivity and contribution of a research team is generally evaluated on a team basis rather than individually, a lot of the research in the area of knowledge building examines the progress and development of the entire community, and this is also the focus of research for the present study. Individual students' advances in understanding or how that relates to group or community advancement in a knowledge building context is not the focus here.

There has been published work in recent years documenting even young children's ability to build knowledge and pedagogical strategies to foster the socio-cognitive dynamics of knowledge building (Scardamalia and Bereiter 1994). In recent years, knowledge building as a pedagogical approach has gained in popularity, with the emergence of some sizeable networks of teachers using the Knowledge Forum® platform to support the learning and teaching process, and to implement the knowledge building approach. Examples of such networks include the TACT project in Canada (http://www.telelearning-pds.org/tlpds/) and the KBTN project in Hong Kong (http://kbtn.cite.hku.hk/). The emergence of such networks of teachers and classrooms provides a unique opportunity for empirical explorations that span across different classrooms to identify trends and regularities in the growth and development in knowledge building in a community of learners (including the teacher who is designing and facilitating the educational process) in classroom contexts. For example, is there a developmental trajectory for classrooms learning to engage in knowledge building? Would classrooms in different cultural and educational contexts follow different developmental trajectories? These are some of the questions that would benefit from the analysis of large corpora of online discourse data that would contribute to theoretical and pedagogical developments in knowledge building as an educational approach.

The problem we want to tackle in this study is to develop a methodology that can take advantage of the increasing collection of online corpora in Knowledge Forum® for the incremental establishment of profiles and models of knowledge building behavior and interactions that reflect different levels of productivity in knowledge building. Studies of knowledge building in educational settings have mainly focused on interactions and outcomes within a designed setting, often in the form of a single classroom (for example, Rahikainen et al. 2001; Lee et al. 2005). Various indicators to gauge the process and outcomes of knowledge building in classroom settings have been reported in the literature. In this study, we build on previous work in the area to identify a number of indicators for exploratory automatic encoding and visualization and implement them on a number of Knowledge Forum corpora collected from primary and secondary classrooms in Hong Kong that engaged their students in knowledge building around the themes of energy crisis and global warming. In this chapter, we begin with a review of some of the popular indicators reported in the literature in studies of knowledge building. This is followed by a brief description of a precursor to this study in which we explored multiple visualizations of a number of machine-generated indicators in conjunction with in-depth qualitative analysis of discourse threads to examine whether such visualizations were helpful in shedding light on the quality of the discussion from the perspective of knowledge building. The methods and findings from the current study and their methodological implications are then presented and discussed.

3.2 Indicators for Knowledge Building Advancement in the Literature

Knowledge building is a collaborative process whereby a community intentionally advances the collective state of knowledge through a knowledge transforming discourse. Knowledge building is not something that "naturally" happens but has to be fostered and developed intentionally as well. Knowledge building theory treats ideas, theories and hypotheses as intellectual artifacts and objects of inquiry that can be discussed, scrutinized, improved and put into new uses as participants engage in progressive inquiry. Knowledge building focuses on developing progressively better understanding and contribution to better collective knowledge as a result of the collaborative intentional effort, and learning takes place as a by-product of the process (Scardamalia et al. 1994). Discourse is hence central to knowledge building. Another corollary of the theoretical underpinning of the knowledge building paradigm is the preference for not putting students into groups, but rather, to encourage learners to read widely the contributions (in the form of discussions) of all members; and to not confine their responses to discussions only of certain group members but to respond if they think they can make a contribution to the discussion.

Twelve knowledge building principles (be referred to as KB principles for short) have been proposed by Scardamalia (2002) as the key characteristics of knowledge building behavior. These principles include democratizing knowledge, idea diversity, epistemic agency, improvable ideas, constructive uses of authoritative sources, real ideas and authentic problems and rise above. Many of the indicators used in analyzing knowledge building discourse are derived from these 12 KB principles. For example, Lee et al. (2006) and Zhang et al. (2007) assessed students' knowledge building outcomes through in-depth qualitative analysis of the discussion notes as well as reflection or portfolio notes on the basis of these principles. Van Aalst et al. (2002) used four criteria derived from the 12 KB principles to assess students' progressive development based on their knowledge building discourse and found evidence for a hierarchy in the students' ability to demonstrate the KB principles in the course of their work. Law and Wong (2003) devised a rubric for rating students' progress along the 12 KB principles based on observations of students' participation in online discussion as well as the quality and level of engagement reflected in the content of their discourse. However, all of these methods are not easily scalable for use on multiple large sets of discourse data as these depend on human judgment on the discussion content and are difficult to automate through machine coding.

3.2.1 ATK Indices as Indicators for Knowledge Building

The Knowledge Forum® platform has been designed as a collaboration space to provide various technological affordances to support the socio-metacognitive dynamic of knowledge building (Scardamalia 2002) work of knowledge building communities. The collaboration space is structured in the form of views, where participants post their ideas in the form of notes. Postings are referred to as notes rather than messages as these are to be considered as intentional inquiry efforts. Notes can be edited multiple times just as in the case of academic writing. Note authors can use a number of features such as *scaffolds*, *keywords*, *references* and *annotations* to support the collaborative knowledge building process. Participants can build-onto existing notes to advance the inquiry, resulting in thread like discussion structures. Views can be used to spatially structure and organize discussion around different themes and/or different groups of participants. New views can also be introduced for participants to move into the next phase of discussion.

The Analytic Toolkit (ATK) (Burtis 1998) is a suite of online analysis tools created to report summary statistics on activities in Knowledge Forum®. It provides statistics on the level of participation of selected members within a time span, such as which view(s) a member has been working in, number of notes posted in selected views, the number of seed notes or build-on notes created, and the percentage of notes in the selected view(s) that have been read. It also provides statistics on the use of specific knowledge building support features such as the use of scaffolds, keywords, reference and annotations. These latter statistics reflect more intentional and sophisticated use of the platform features to advance understanding and knowledge building. Among these, scaffold use has been the most popularly adopted among the many features available in Knowledge Forum®. The most frequently used scaffolds include *My Theory, I want to understand, This theory cannot explain, A better theory, New information, My opinion and Putting our knowledge together.* Scaffolds can be custom designed by discourse managers (often the teachers) to suit the curriculum objectives. Participants are free to decide which and how many scaffolds to use in the note-writing process.

Use of scaffolds is considered an important knowledge building indicator in the literature, reflecting participants' socio-cognitive (Scardamalia 2002) and socio-metacognitive (Law 2005) engagements. Tan et al. (2005) use the frequencies of use of scaffolds as indicators of students' developing scientific inquiry skills through the collaboration process. Laferriere and Allaire (2005) investigated the use of a set of scaffolds created on the basis of the 12 KB principles in a pre-service teacher education program over a 3-year period. Based on quantitative analysis of the online activities and content analysis of the online discourse in Knowledge Forum®, the study found that over time, there was an increase in the frequency and diversity of the scaffolds used as well as an improvement in the appropriateness of the choice of scaffolds selected with respect to the content and focus of the notes contributed by the student teachers, indicating an advancement in their socio-metacognitive development in knowledge building. Law et al. (2008) report that in a knowledge building discourse involving grades seven and nine students on global warming, notes in more sustained discussion threads with more than four levels of build-on notes included the use of a larger number and more appropriate sets of scaffolds for the foci and nature of the note contents than those in threads that are less developed. Based on the results from factor analysis on ATK measures, Lee et al. (2006) found that of the two factors obtained, the one comprising the numbers of notes created, notes read, scaffolds used and notes revised (to be referred to as ATK Inquiry Index) can explain a high percentage of the variance in students' achievement.

3.2.2 Questioning and Level of Inquiry

Students' self-generated questions are important starting points for students' inquiry (Scardamalia and Bereiter 1991, 1992). Hence, the presence and nature of questions in the online discourse has been used as indicators for the level of inquiry in knowledge building. Hakkarainen (1998) classified questions into factual and explanation oriented questions. The former are concerned with the search for some basic information while the latter look for answers that reveal deeper levels of understanding and are hence more likely to lead to productive knowledge building. Fact-oriented questions include those concerned with ‘which’, ‘where’ and ‘when’, while questions concerned with ‘why’ and ‘how’ are considered as explanation-oriented questions. In this study, the quantity and nature of the questions included in the students’ discourse on the Knowledge Forum are used as indicators for students’ knowledge building engagement.

Questions are important not only at the beginning stage of a knowledge building discourse to spark off the inquiry. The progressive generation of subordinate questions is also crucial to the advancement of inquiry and knowledge building (Hakkarainen and Sintonen 2002). Hakkarainen (1998) identified three types of inquiry patterns according to the generation of subordinate questions in the process: (1) truncated inquiry, which ends before any subordinate questions are generated; (2) extensive inquiry, which contains several unrelated principal questions, and each with only a few subordinate questions generated; and (3) intensive inquiry, characterized by the in-depth exploration of a topic through the generation of a series of subordinate questions. In the present study, we also examine the patterns of inquiry in the discourses we analyze.

The level of students’ inquiry is also reflected through the level of explanation found in their discourse. Hakkarainen (1998) classified students’ discourse into five levels of explanation: (1) isolated facts; (2) partially organized facts; (3) well-organized facts; (4) partial explanation; and (5) explanation. In a more recent study, Zhang et al. (2007) developed a similar framework for analyzing the depth of inquiry in knowledge building with four levels of “epistemic complexity of ideas”: (1) unelaborated facts; (2) elaborated facts; (3) unelaborated explanations; and (4) elaborated explanations. The definitions of the higher levels of explanation in both of these two classification frameworks assume that there is a reference explanation that reflects the best current knowledge about the domain of inquiry, which is often the case in the area of science and technology, but not necessarily appropriate for inquiries in the humanities and arts areas.

Veldhuis-Diermanse (2002) adopted the Structure of Observed Learning Outcomes (SOLO) taxonomy developed by Biggs and Collis (1982) to analyze the quality of students’ constructed knowledge in a CSCL context. The SOLO taxonomy was developed to measure students’ learning outcome in different subject areas and across different contexts. It comprises five levels: (1) prestructural, when a student may not have understood the task at all (2) unistructural, when one aspect of the task is identified, (3) multistructural, when several aspects of the task are identified, (4) relational, when several aspects are identified and integrated into a coherent whole, and (5) extended abstract, when a coherent set of concepts is generalized to a higher level of abstraction. It can be seen that the definition for the lower levels of outcome according to the SOLO taxonomy is very similar to the schemes used by Hakkarainen (1998) and Zhang et al. (2007). On the other hand, the definitions for the higher SOLO levels do not rely on having access to what counts as a complete or more appropriate explanation. Furthermore, at the highest level – extended abstract – the focus is on theorizing, generalizing, hypothesizing, reflecting, and concluding, which is compatible with the focus in knowledge building. We adopted the SOLO taxonomy as indicators for depth of inquiry in another study which we will report briefly in the next section. However, as the coding of online discourse according to the SOLO taxonomy requires rather sophisticated human judgment which is not easily automated through machine analysis, this has not been included in the set of indicators used in the present study.

3.2.3 Argumentation as Indicative of Inquiry and Engagement in Discourse

The theory of knowledge building is grounded on Popper's (1972) theory that theories and ideas are World three artifacts and that human knowledge advances through the improvement of these World three artifacts. Drawing on the practice in scientific communities, Popper considered the argumentative function of human language as the main instrument for the growth of knowledge. Collaborative argumentation has also been found to be contributive to deepening understanding in CSCL discourse (Andriessen et al. 2003). On the other hand, argumentation as debate and persuasion does not bring about knowledge building (Scardamalia and Bereiter 2006), and indicators of knowledge building grounded on argumentation are relatively rare. It is our view that it is possible to differentiate these two forms of argumentation and that analyzing inquiry-oriented discourse from the perspective of argumentation can shed light on the nature and quality of the knowledge building taking place through the discourse. There are studies demonstrating that argumentation has the potential to produce critical inquiry towards knowledge building goals if it is properly used (see e.g. Andriessen et al. 2003; Andriessen and Coirier 1999). In a study of the online interactions between two classes of students, one having several years of prior experience in knowledge building and the other being novice to the Knowledge Forum® platform and the pedagogical approach, Lai and Law (2006) found that students from the former class were more able to problematize taken for granted ideas and concepts, thereby contributing to the generation of dissonance and deepening understanding. They also argue on that basis that analyses of discourse from an argumentation perspective can more readily highlight the presence of problematizing moves and negotiation of meaning, both of which are contributive to knowledge building.

There are distinctive linguistic markers associated with argumentation that can be used to analyze computer-mediated collaborative learning discourse (Saab et al. 2005). Although these discourse markers do not contribute a great deal to the semantic content of the utterances in which they occur, they do determine how the semantic content of such utterances are interpreted. Studies of CSCL discourse find that discourse markers indicative of various kinds of argumentation acts are often present when the learners are deeply engaged in expressing their ideas to advance their knowledge and understanding through the discourse (van Boxtel et al. 2000).

3.3 A Study to Identify Features of More Productive Discourse Threads

Based on the rich array of indicators reviewed above, we conducted a number of studies to explore whether we can identify a set of indicators and tools that can be used to provide a holistic representation of discourse threads that can shed light on how far an individual thread has been productive in terms of the knowledge building efforts it reflects that triangulates well with in-depth qualitative encoding involving expert human judgment. Here in this section, we present a brief summary of a recent study (Law et al. 2008) that provides the empirical basis for the design of the work reported in this chapter.

Law et al. (2008) report on a study that analyzed the collaborative online discourse of students studying in grades 7 – 9 from two secondary schools in Hong Kong. The topic of inquiry was global warming, and the teachers were consciously adopting the knowledge building approach (Scardamalia and Bereiter 2003) and Knowledge Forum® was used as the platform for inquiry and collaboration among the students. The collaboration lasted about 6 weeks, resulting in 194 notes written over this period. Careful qualitative coding of each note using the SOLO taxonomy (Biggs and Collis 1982; Biggs 1999) was carried out as an indicator of the quality of the discussion at note level, and the mean SOLO score for notes in a thread were also used as an indicator of the level of inquiry at the thread level. The ATK Inquiry Index (Lee et al. 2006), argumentation speech acts (including the kind of questions asked), and scaffolds used were used as quality indicators for the inquiry process. The research team found that most of the discussions and resource materials available in the public domain on global warming focused on three distinctive areas of concern: scientific evidence and explanations for global warming, policies and strategies as well as challenges to deal with global warming from a sociopolitical perspective, and personal action to reduce global warming. The Law et al. (2008) study also explored whether the depth and quality of inquiry were related to the content focus of the inquiry threads.

Three sets of tools were used in this study. The Analysis Tool Kit (ATK) (Burtis 1998) was used to compute the ATK Inquiry Index for each student based on the totality of their participation (both reading and writing) in the online discourse. The Visual Intelligent Content Analyzer (VINCA, a tool set jointly developed by the Institute of Knowledge Science and Engineering at Beijing Normal University and the Centre for Information Technology in Education at the University of Hong Kong) (Law et al. 2007) capable of processing textual records in both English and Chinese was used to (1) code each note in the discourse for the kind of scaffolds used and the kind of argumentation speech act present through matching of specified text patterns of argumentation discourse markers in the discourse, and (2) compute the semantic closeness of the content of the notes in each extended discussion thread with each of three sets of text selected from public media with the three different content foci identified above. The third tool used was the Knowledge Space Visualizer (KSV) developed by Teplovs (Teplovs and Scardamalia 2007) to visualize the discourse development based on the coding generated from VINCA. The KSV is a program tailored to reveal discourse dynamics arising from the use of Knowledge Forum® by displaying hidden discourse relationships and interactions between notes in graphical representations.

Examining the various indices at the individual student level, the study found no relationship between the ATK inquiry index and the mean SOLO level of the notes written by a student. This indicates that the ATK inquiry index, which is primarily computed from participatory statistics, bears no inherent relationship with the quality of a student's written contribution to the discourse (SOLO level). There was also no relationship found between the ATK inquiry index and the number and pattern of argument markers used by a student or the number and nature of question markers (i.e. whether the discourse markers indicate the presence of factual or explanation-oriented questions). On the other hand, students with higher ATK inquiry index tend to use more scaffolds, indicating a higher familiarity with the use of scaffolds for knowledge building purposes. However, students familiar with scaffolds as a Knowledge Forum® feature do not necessarily show deeper levels of engagement or higher quality contribution to the discourse.

Analyses of the indices at the thread level provide more interesting findings. Firstly, it was found that singletons, doublets and notes in threads with fewer than five notes had very few or no question markers and very low presence of scaffold use, although these had similar numbers of argumentation markers per note compared to those in the longer threads. The SOLO levels of the notes in singletons and doublets were very low, mostly at unistructural or multistructural levels. For threads with four or more levels of build-on notes, there appears to be an identifiable pattern of codes in their sequential development. Most of these more "productive" threads (in the sense of being more sustained discussions) started with notes at the multistructural level, progressing to its peak (relational or extended abstract) towards the middle of the thread, and reverting to multistructural or unistructural level when the thread ended. It was observed that notes coded at high SOLO levels (relational or extended abstract) were more likely to contain scaffolds such as "My Theory", "A Better Theory", and "Opinion", as well as argumentative markers for claims, contrasts and rebuttals (Law et al. 2009). Notes coded at low SOLO levels (unistructural or multistructural) had low use of scaffolds and argumentative markers. Further, a difference in the pattern of argumentation markers was found in notes belonging to extended threads with five or more notes compared to the others. These notes, particularly for those found in the first or second notes in the extended threads, were much more likely to contain markers indicating the use of contrast, question and rebuttal in their content. The study also found that of the eight extended threads found in the discourse, only one had a science-related content focus while the others were mainly focused on socio-political concerns.

In summary, this study indicates that threads shorter than five notes are unlikely to demonstrate deep levels of knowledge and inquiry and that scaffold use and the presence of question markers and argumentation markers associated with exploration of ideas such as contrasts and rebuttals seem to differentiate more productive discussion threads from others. The content focus of the discussion thread also seems to matter. The findings from this study inform the empirical work that is presented in the rest of this chapter. The corpus analyzed in the Law et al. (2008) study briefly reported here is also included as one of the seven corpora analyzed in the present study.

3.4 Methods and Tools Used in this Study

In this study, in addition to examining basic participation statistics, we decide to explore the quality and progress in knowledge building as reflected by discourse data from three perspectives based on the literature and our earlier work: Knowledge Forum® scaffolds used, argumentation and questioning speech acts involved and the subject matter content of the discourse. While it is relatively easy to identify the scaffolds since these are built-in to the Knowledge Forum® platform, the other two analyses foci are more complex. In order to cope with the large amount of discourse data included in the seven corpora (see Table 3.1 for some basic statistics of the analyzed corpora), machine-supported coding is employed in this study.

There are two common approaches found in current work on automatic coding of discourse data. One approach is to use machine-learning algorithms to generate increasingly more reliable coding based on manually coded corpora (e.g. Soller 2004; Goodman et al. 2005; Gweon et al. this volume). The coding (or categorization) targeted in these studies tend to be more complex, global ones such as whether the collaboration was effective or whether there is knowledge building taking place, and is hence less suitable for the kind of rather specific coding at a note level that we want to accomplish in this study. Another approach is to use keywords, key phrases or text patterns to identify specific discourse characteristics (e.g. Pennebaker et al. 2007; Law et al. 2007; Erkens and Janssen 2008). While this latter approach has been criticized as relatively simplistic and may not be able to shed light on more complex processes (e.g. Rosé et al. 2008), it is more suited for the coding of dialogue acts. To date, there is a rich body of computational linguistic research in this area (Heeman et al. 1998; Stolcke et al. 2000; Janssen et al. 2007a; Erkens and Janssen 2008). In this study, we have used matching of text patterns for the coding of argumentative and questioning speech acts.

Identifying the subject matter content focus of discourse notes is much more complicated compared to the coding of speech acts, and different methods may be necessary for different purposes. One common approach to automated semantic analysis is to compute the similarity among pieces of text. Latent semantic analysis utilizes a form of mathematical matrix decomposition technique to determine similarity in the meaning of words and passages through analyses of large text

Table 3.1 Basic statistics about the corpora analyzed in this study and some background information about the classrooms that generated them

Corpus ID and label	Teacher	year	Period (days)	#Views	Total #notes	#Auth.	#notes/ author	Av. #notes/ day	# Threads	#Notes/ thread	#Thread >4 notes	%Thread >4	Av. #notes for threads >4 notes	#Notes in longest thread
1. X6A	A	2006	21	2 views	201	13	15.5	9.6	31	6.5	14	45.2	11.6	34
2. X6B	A	2006	28	2 views	169	15	11.3	6.0	17	9.9	7	41.2	21.3	53
3. X6C	B	2006	35	1 views	139	15	9.3	4.0	5	27.8	4	80.0	34.5	106
4. X6_07	A	2007	50	4 views	212	26	8.2	4.2	53	4.0	19	35.8	8.0	14
5. Y7C	C	2007	78	3 views	180	38	4.7	2.3	79	2.3	8	10.1	8.8	17
6. Y78Z9	C+D	2007	43	1 view	194	49	4.0	4.5	37	5.2	12	32.4	12.9	28
7. X56QIP	C+E	2009	82	4 views	1,275	63	20.2	15.5	224	5.7	38	17.0	26.2	182

corpora (Landauer et al. 1998). The semantic closeness (i.e. text similarity) between pairs of texts can also be computed based on the Vector Space Model (Salton et al. 1975). This approach is useful in discovering the similarity between text passages, for example in identifying the content similarity between discourse notes and identified expert entries (Law et al. 2008). However, in the present study, we need a finer granularity in the identification of content focus – we want to identify the occurrence each time a specific content is mentioned in the discourse. We have hence used text matching of keywords and keyword patterns for this purpose, as described in some detail later in this section.

The specific indicators and tools for their identification are described below.

3.4.1 Scaffold Supports Used

The analysis identifies the Knowledge Forum® scaffold supports commonly used by teachers. The kinds of support and associated scaffolds coded are as follows:

Support type	Scaffolds coded
Information	New Information, Source of information
Query	I need to understand
Proposing Theory	My theory
Opinion	Opinion, Different Opinion
Elaboration	Reasons, Example, Evidence, Elaboration
Theory Exploration	This theory cannot explain, A better theory
Rise Above/Summarize	Putting our knowledge together, conclusion

3.4.2 Discourse Acts Related to Argumentation and Questioning

Argumentative markers (van Boxtel et al. 2000) can be used to identify some common speech acts that are often associated with inquiry. The following speech acts and the related specific discourse markers used in the analysis are listed below:

Speech act	Discourse markers coded
Claim	I think, I agree, we should
Disagreement	I don't think, I didn't think, I do not think, I don't agree, I do not agree, I didn't agree, we shouldn't, we should not
Reason/Elaboration	moreover, such as, because, since
Condition	if
Contrast	but, though, although, however, even, otherwise
Consequence	then, thus, so, therefore
Explanation questions	how, why
Factual questions	what, is there, are there, where, who, whom

3.4.3 Domain-Specific Topics and Associated Keywords and Text Patterns

The identification of relevant domain-specific keyword patterns provides researchers an efficient means to capture and examine the content focus of a discussion. Based on concordanced listings of text patterns around high frequency keywords extracted from the notes posted in Knowledge Forum®, a list of domain related keywords and word patterns are categorized into seven topic areas related to ecology, energy and global warming. These are then used by VINCA to code the presence of these content topics for use in the later analysis and visualization. The content topics and a brief description of the associated keywords and text patterns are listed as follows:

Content topics	Associated keywords and text patterns that indicate
Ecology	Food chain, food web or habitats
Field investigation	Parameters or measurements as evidence, e.g. conductivity
Environmental problems	Forms of pollution and their consequences
Energy sources	Different energy sources, e.g. solar, tidal, wind, etc.
Mechanisms for generating electricity	Use of technical words to describe mechanisms, e.g. reactor, electrolyte, chain reaction, etc.
Sociopolitical issues	Sociopolitical impacts and strategies, e.g. economy, Kyoto Protocol, etc.
Personal action	Actions to reduce global warming, e.g. reduce, recycle

Clearly, coding for the presence of content topics using the keywords and patterns listed is a very crude approach and some relevant discussions maybe omitted or not coded properly if the vocabulary used is not associated with a high frequency word picked up from the initial keyword search. VINCA (Law et al. 2007) was used to compute the basic statistical counts of interactions and to code the discourse corpora according to the sets of indicators specified above.

3.4.4 Information Visualization

Information visualization is the mapping of multiple forms and sources of data in visual form, through the use of computer supported, interactive, visual representations of abstract data to amplify cognition by taking advantage of the way human perception operates (Card et al. 1999). There are two main categories of visualization designs – linear and non-linear – that CSCL researchers have used to support sense-making of the many different kinds of analysis conducted on discourse data to shed insight on processes and development at individual and group levels. Linear designs organize multiple streams of information or coding in sequence, mostly chronologically. Many visualization designs belong to this category, such as contingency graphs (Suthers and Medina, this volume), CORDTRA diagrams (Hmelo-Silver et al. this volume), Tatiana (Dyke et al. 2009) and Bobinette on the Calico platform (Giguet et al. 2009). Linear visualization designs have two apparent advantages. First, they present various layers of discourse dynamics in chronological order, making it easier for researchers to look for patterns of developmental trajectories across layers. Second, a large diversity of information types (e.g. 40 in CORDTRA diagrams) can be incorporated in one graph and still allow relatively easy sense making. Non-linear visualizations on the other hand tap on the visual-spatial sense of human cognition to provide meaningful snapshots of CSCL discourses. Social network analysis diagrams (Scott 2000), the Participation Tool (PT) developed by Janssen et al. (Janssen et al. 2007b) and Teplovs' (2008) Knowledge Space Visualizer belong to this category.

In this study, a linear visualizer, the Bobinette tool is used to visualize the coded discourses in thread structures over time. Bobinette is developed by Pierre Lecavelier and Emmanuel Giguet from an original idea of Benjamin Huynh Kim Bang and Eric Bruillard, and is part of the Calico project, Calico is a development project located at CNRS, Université de Caen Basse-Normandie France, ENSICAEN, to design and create a set of tools to study and monitor computer mediated communication (CMC). Users can upload, view, study and share CMC objects like data or content analysis coding schemes in Calico.

The Bobinette diagrams provide chronological visualization of the presence of up to eight different specified codes (labeled as themes in the software) in threaded discourse entries. For example, Fig. 3.1 in Sect. 3.5 is a Bobinette display of the presence of argumentation and question markers in some of the earlier threads in one of the corpus analyzed in this study. The second row of text provides a legend for the color coding used for the representation of the different discourse markers. Each row represents a thread, with the threads sequenced chronologically according to the first note in the thread. The display cannot show the thread structure, and only displays the notes contributed on the same day in the same column (each note is represented as a circle). The presence of color-coded squares following a circle indicates the presence of the specific discourse markers in that note, and the number within a square indicates the number of times that specific marker is found in the content of the note. For example, there are three circles (labeled a, b and c) in the first thread under the column dated 09/01, indicating that three notes were added to this thread on this date. Note a contains one discourse marker indicating the presence of a factual question. Note b contains one argumentative discourse marker for consequence, indicating that one of the markers – then, thus, so or thus – was used once in this note. Note c contains two different discourse markers, one for consequence and the other for a factual question. Figure 3.2 is another Bobinette display. The first note in the first thread displayed contains four discourse markers, three of which are expressions of contrast (i.e. words within the list for contrast – but, though, although, however, even, otherwise) and one is a marker for a factual question.

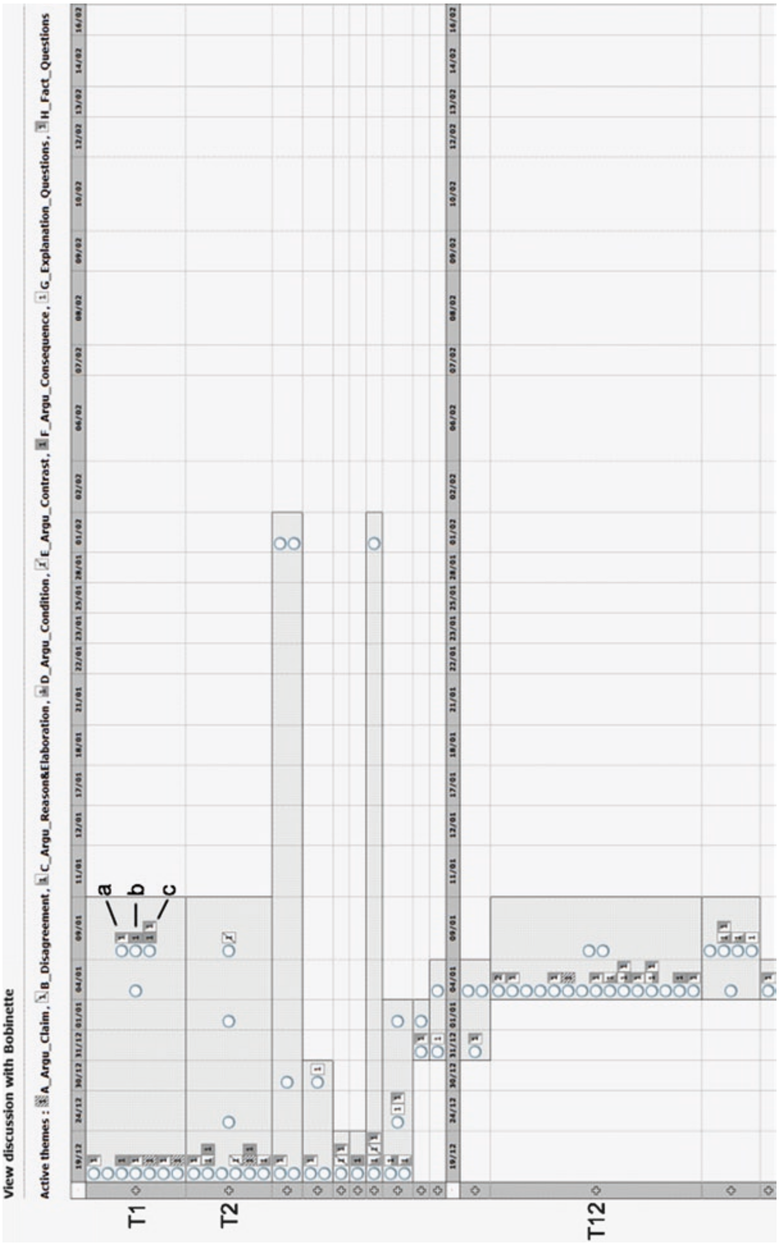

Fig. 3.1 A section of the Bobinette display of the prsence of argumentation and question markers in the first 14 threads in the Y7C corpus. Each circle represents one note contributed to the view. Each patterned box following a note represents the presence of that particular type of argumentation or question marker in that note, and the number within the box indicates the number of that kind of marker within that note

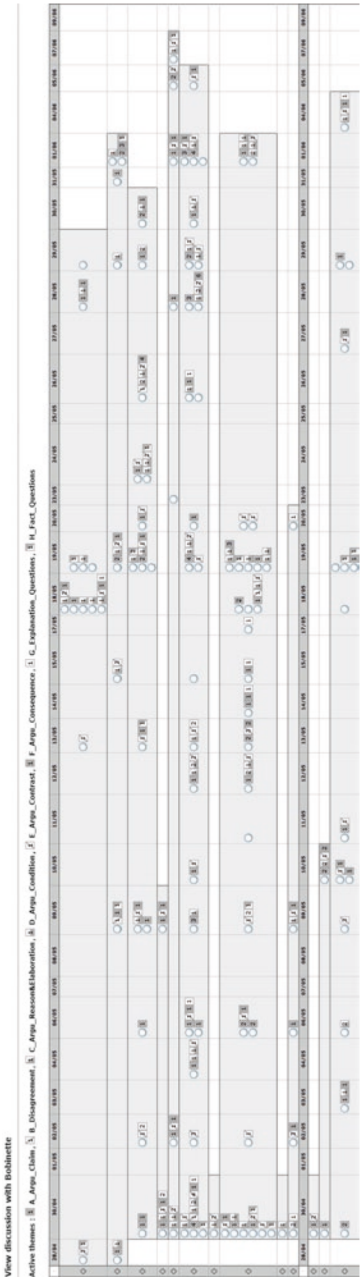

Fig. 3.2 A section of the Bobinette display of the presence of argumentation and question markers in the first 13 threads in the Y78Z9 corpus. Each circle represents one note contributed to the view. Each patterned box following a note represents the presence of that particular type of argumentation or question marker in that note, and the number within the box indicates the number of that kind of marker within that note

3.5 The Context and Basic Quantitative Information of the Discourse Corpora Analyzed

In order that the current study can build on our previous work, and in particular to be able to make use of the content coding schemes developed earlier, we have selected the discussion corpora on Knowledge Forum® generated from seven "classrooms" that engaged students in knowledge building around energy crisis and/or global warming. These corpora were collected at different times over the past few years. Some background information about the classroom contexts and basic statistics about these corpora are presented in Table 3.1.

Three of the corpora came from the work of three grade six classrooms (6A, 6B and 6°C) in School X (primary) done during the 2005–2006 academic year. The work in classes X6A and X6B were led by teacher A and the work in class X6C was led by teacher B. Teacher A had experimented with using Knowledge Forum® for a short period of time in the previous school year, while this is the first experience for Teacher B to adopt the knowledge building approach in his teaching. Teacher A continued to work with another class of grade six students (class X6_07) in the following school year to knowledge build around the theme of global warming and energy crisis. During that same academic year, from early January to the beginning of March, teacher C in School Y (primary) joined the Knowledge Building Teacher Network and she experimented with engaging students in one of her grade seven classes, Y7C, in knowledge building around the theme of energy crisis. Later, she found out that another teacher in the network, Teacher D in school Z, had been working with a class of grade nine students on knowledge building around the same topic, and that the topic of global warming came up as a major theme during the process. Teachers C and D then decided to try getting students in the two schools to work together on the same discussion view on Knowledge Forum®. For School Z, all students from class Z9 would continue with the knowledge building work, while in school Y Teacher C decided to conduct this part as an extra-curricular activity with a small number of students from several classes in grades seven and eight. The corpus generated during this collaboration period is labeled Y78Z9 in Table 3.1. This is the same corpus that was analyzed in the study reported earlier in Sect. 3.2. Teacher C then joined school X in the school year 2008–2009, and from March to May 2009, she conducted whole curriculum topics on ecology, global warming and energy using the knowledge building approach. These students engaged in discussing the topic of global warming with a class of grade six students in Quebec as part of an international knowledge building program.

The above description clearly shows that the experience and understanding of the teachers and students involved vary great across the seven sets of knowledge building corpora analyzed in this study. Hence it is very likely that the quality and depth of inquiry mediated through these corpora also differ significantly, which is in fact the case as will be demonstrated later through selected examples. Further, if the depth of inquiry through the online discourse is to increase with the facilitating teacher's experience with the knowledge building approach and the Knowledge Forum® environment, then the more recent corpora are more likely to reflect deeper levels of inquiry compared to the earlier ones. In the rest of this chapter, we will examine and compare the different pictures of the quality of the discourses as reflected by the different sets of selected indicators, finishing with a discussion of the implication of the findings in relation to the identification of indicators that can be used to build up profiles of knowledge building behavior and trajectories of development.

A number of the parameters in Table 3.1 are potentially useful indicators of student engagement in knowledge building. The average number of notes per author and notes per day are purely quantitative reflections of how far students were engaged in the writing process and do not reflect the quality of the engagement. Hence these quantitative indicators per se cannot be taken as quality indicators for knowledge building. On the other hand, indicators related to the length of discussion threads such as the average length of threads do reflect to some extent the potential for the discourse to be a productive one as the depth of inquiry in a short thread is necessarily limited. Hence the percentage of threads with more than four notes is a potentially more informative indicator. Using this criterion, the depth of inquiry in the two corpora with the lowest mean number of notes per thread and the lowest mean number of notes in threads longer than four notes, X6_07 and Y7C, are probably doubtful. This will be further explored in a later section.

One prominent observation from Table 3.1 is the lack of any perceptible developmental trend in terms of the quantitative indicators of depth of inquiry (including the average number of notes per thread, the percentage of notes in threads longer than four notes, the average number of notes in threads having more than four notes and the number of notes in the longest thread) in the analyzed corpora. In fact, the statistic for the third corpus generated by a class of students new to knowledge building and Knowledge Forum® and facilitated by a teacher who is likewise novice to knowledge building, X6C, is somewhat surprising. It has only five threads and 106 out of the total of 139 notes were found in one thread. In revisiting the discourse views for this class in Knowledge Forum®, we find evidence that the "long thread" is probably an indication of the lack of understanding of how to introduce new threads to start a discussion on a new topic or problem. Forty-two of the 139 notes in the corpus were written on the first two days, and these were all linked to the first note as one single "threaded discussion". Three new threads were started subsequently but most students still preferred to enter their notes as "build-ons" to the very first thread. The last "thread" consisted of only one note added 3 weeks after the second last note was added to the first thread. Hence quantitative indicators of depth of inquiry such as the average thread length and the percentage of notes in long threads are not likely to be good indicators of depth of inquiry when used as standalone indicators by themselves.

From this analysis, we have clear evidence that the experience and expertise of the teacher in relation to facilitating knowledge building matters. Something as basic as the concept of "discussion threads" and when a new discussion thread should be started cannot be taken for granted. This has implications for teacher professional development in CSCL. Another implication from this analysis is that researchers need to be very cautious in interpreting mechanized analysis results at face value. A "reality" check through a review of the actual data set and the discourse content is necessary to ensure the validity of the interpretation.

3.6 Comparing Corpora Using Discourse Markers

Next, we examine the presence of argumentation and question markers in the discourse. As discussed earlier, it is expected that a higher concentration of argumentation markers is likely to indicate a stronger socio-metacognitive engagement and inquiry orientation in the discussion, and hence a higher probability for the discourse to be productive in terms of knowledge building. The statistics in Table 3.2 indicate that the concentration of argumentation markers in the first five sets of corpora, X6A, X6B, X6C, X6_07 and Y7C are rather similar and generally lower than the same statistics for the other two corpora. This seems to indicate that the notes in the two later corpora had a stronger orientation towards the exploration of ideas. In particular, except for argumentation markers associated with reason and elaboration which had the highest concentration in the X56QIP and X6_07 corpora, the concentration of all the other five argumentation markers were highest in the Y78Z9 corpus. It is true that students and teachers involved with the Y78Z9 and X56QIP corpora were more experienced in knowledge building. However, it is also important to note that students involved in the Y78Z9 corpus were the oldest in age and hence may be more able to communicate and express their ideas linguistically than the other, younger students.

When the markers for explanation questions and factual questions are examined, a somewhat different trend is observed. While the first three corpora, X6A, X6B and X6C still had the lowest concentration of question markers per note, X6_07 and Y7C had respectively the highest concentration of explanatory and factual questions, and both being well above the concentration of questions even in the corpora Y78Z9 and X56QIP. This seems to indicate that students participating in these two corpora were interested in the problems being discussed and had identified questions for inquiry. However, the percentage of notes without either argumentation or question markers is still relatively high compared to the two later corpora Y78Z9 and X56QIP. One possible interpretation of this figure is that even though students associated with the corpora X6_07 and Y7C had identified questions they wanted to inquire, they were less able to construct a knowledge building discourse to advance the inquiry. This hypothesis is corroborated by the statistics in Table 3.1, which reveals that X6_07 and Y7C had the lowest values for all the quantitative indicators for sustained inquiry: the mean number of notes per thread, the mean number of notes for threads with more than four notes the mean number of notes in the longest thread.

Since Y7C and Y78Z9 also had some relatively extended threads, we try to examine the Bobinette displays of the notes from the two corpora to see if we can find any

Table 3.2 The total number and mean number per note of the various kinds of argumentation and question markers present in the analyzed corpora

Corpus ID and label	Argu_Claim (mean # per note)	Argu_Dis-agreement (mean # per note)	Argu_Condition (mean # per note)	Argu_Reason & Elaboration (mean # per note)	Argu_Contrast (mean # per note)	Argu_Con-sequence (mean # per note)	Total of argumentative discourse markers	Explanation Questions (mean # per note)	Fact Questions (mean # per note)	% of notes without any arg_marker OR qn_marker
1. X6A	41 (0.20)	7 (0.03)	23 (0.11)	55 (0.27)	54 (0.27)	35 (0.17)	215	11 (0.05)	17 (0.08)	60 (0.30)
2. X6B	37 (0.22)	2 (0.01)	32 (0.19)	37 (0.22)	45 (0.27)	34 (0.20)	187	16 (0.09)	16 (0.09)	60 (0.36)
3. X6C	23 (0.17)	2 (0.01)	18 (0.13)	24 (0.17)	46 (0.33)	17 (0.12)	130	15 (0.11)	6 (0.04)	59 (0.42)
4. X6_07	32 (0.15)	3 (0.01)	35 (0.17)	65 (0.31)	68 (0.32)	54 (0.25)	257	52 (0.25)	43 (0.20)	52 (0.25)
5. Y7C	21 (0.12)	1 (0.01)	22 (0.12)	20 (0.11)	25 (0.14)	24 (0.13)	113	18 (0.10)	38 (0.21)	84 (0.47)
6. Y78Z9	75 (0.39)	13 (0.07)	53 (0.27)	57 (0.29)	82 (0.42)	62 (0.32)	342	21 (0.11)	25 (0.13)	21 (0.11)
7. X56QIP	431 (0.34)	23 (0.02)	333 (0.26)	391 (0.31)	334 (0.26)	309 (0.24)	1,821	218 (0.17)	212 (0.17)	254 (0.20)

structural difference between them. Segments of the two Bobinette displays generated using Calico (Giguet et al. 2009) are presented in Figs. 3.1 and 3.2.

There are three threads in Fig. 3.1 that have nine or more notes: threads T1, T2 and T12. However, examining the distribution in terms of time of entry of those notes, it is hardly appropriate to refer to any of these threads as "long" threads indicative of extended inquiry. The patterns of participation in all the discussion threads irrespective of the thread length are very similar – most of the notes were entered on 1 or 2 days and the threads ended soon after. It is very likely that most of the notes were entered when students were given an opportunity to work on the Knowledge Forum® platform during class time. Unfortunately, the inquiries that were started in class did not seem to be pursued further by the students. Hence the structure is typical of truncated inquiries as described by Hakkarainen (1998).

Figure 3.2 presents a very different image of what was taking place in the Y78Z9 corpus. First of all, even though there may be many notes entered by students on the same day, they tend to be placed in different threads, indicating that students probably had thought more carefully about whether what they plan to write fit within existing threads or about new topics that warrant the creation of a new discussion thread. More importantly, most of the threads were followed up throughout the course of the 7 weeks during which the discussion took place. Thus the long threads in this corpus have the structural features indicative of extended inquiry that goes beyond a simple count of the number of notes in a thread.

In this part of the analysis, we find that the use of a visualization tool that displays the online interactions sequentially over time is very helpful in providing a perspective to the knowledge building process which would otherwise be lost if we only examine the statistics of the discourse markers.

3.7 Comparing Scaffolds Used in the Discourses

As mentioned earlier, the numbers and kinds of scaffolds used reflect the students' understanding of the processes involved in knowledge building and their ability to identify the nature and focus of the specific contributions made in the notes that they write. Table 3.3 presents the total and mean numbers of each type of scaffold present in the notes of the seven analyzed corpus. From these statistics, it is quite clear that with the exception of Y78Z9 and X56QIP, the use of scaffolds in the other five earlier corpora rarely go beyond the information scaffold. This possibly reflects a relatively simplistic understanding of knowledge building as locating relevant information. On the other hand, in knowledge building, the identification of the problem(s) of understanding is primary and is the starting point of the knowledge building process. Further, knowledge building is not just an accumulation of relevant information, but more importantly, developing a better, more adequate conceptualization of the problem and formulating a theoretical interpretation of what is known. Hence the scaffolds "I need to understand" and "my theory" are more important than information from the knowledge building perspective.

Table 3.3 The total number and mean number per note of the key scaffolds present in the analyzed corpora

Corpus ID and label	Proposing theory (mean # per note)	Theory exploration (mean # per note)	Query (mean# per note)	Rise above/ summarize (mean # per note)	Information (mean # per note)	Opinion (mean # per note)	Elaboration (mean # per note)
1. X6A	0	0	20	0	21	0	0
	(0.00)	(0.00)	(0.10)	(0.00)	(0.10)	(0.00)	(0.00)
2. X6B	3	0	7	0	47	0	0
	(0.02)	(0.00)	(0.04)	(0.00)	(0.28)	(0.00)	(0.00)
3. X6C	0	0	21	0	35	0	0
	(0.00)	(0.00)	(0.15)	(0.00)	(0.25)	(0.00)	(0.00)
4. X6_07	0	0	0	0	83	0	0
	(0.00)	(0.00)	(0.00)	(0.00)	(0.39)	(0.00)	(0.00)
5. Y7C	0	0	4	1	1	0	2
	(0.00)	(0.00)	(0.02)	(0.01)	(0.01)	(0.00)	(0.01)
6. Y78Z9	68	6	24	12	23	40	0
	(0.35)	(0.03)	(0.12)	(0.06)	(0.12)	(0.21)	(0.00)
7. X56QIP	426	71	244	52	108	212	112
	(0.33)	(0.06)	(0.19)	(0.04)	(0.08)	(0.17)	(0.09)

Examining the statistics further, one can see that the scaffold use in Y78Z9 and X56QIP are more sophisticated. In fact the mean number of queries raised in X56QIP is the highest. Further the students involved in these two corpora not only proposed theories during the discourse. There were instances where they engaged in further exploration of theory by using the scaffold "a better theory". They were also differentiating opinion from information or theory, and made efforts to summarize the discussion in order to move the discussion forward using summarizing scaffolds such as "Putting our knowledge together" and "Conclusion". It is also observed that the X56QIP corpus contained the elaboration scaffolds such as "Reasons", "Elaboration" and "Evidence".

While both Y78Z9 and X56QIP demonstrated more varied and sophisticated use of the metacognitive scaffolds in Knowledge Forum®, an examination of the sequences of scaffolds used in the discussion threads using the Bobinette displays generated from Calico (Giguet et al. 2009) seem to indicate a clearer intentional advancement of the discussion through the use of scaffolds in the X56QIP corpus. Figure 3.3 shows the scaffolds using in segments of several selected discussion threads in Y78Z9. The display shows variations in the scaffolds used in the different threads and at different points in the threads, there is not a clear structure or sequence of scaffolds that one can associate with a logical sequence of inquiry moves.

Figure 3.4 shows the Bobinette display of the first two thirds of the notes contained in the first thread in the X56QIP corpus. The thread starts with the formulation of a query in the first three notes. This was followed by some expressions of opinions and elaboration, followed by further formulation of queries and two attempts to summarize the discussion on March 3rd and 4th. This was followed by some notes providing information and further exploration of theory, which seemed to have sparked off another cycle of exploration through notes containing further queries, further expressions of opinion and elaboration.

Again here, we find that the use of a visualization tool that displays the time sequence in the use of scaffolds to be extremely valuable in revealing whether there is evidence of a cyclic process of progressive inquiry. This discourse feature would be lost to our observation if only the use statistics are examined.

3.8 Comparing the Discussion Content of the Discourses

Based on a preliminary reading of the seven Knowledge Forum® corpora included in this study and the previous detail analysis of the Y78Z9 corpus reported in Law et al. (2008), we find that students' discussions mainly centre around several major themes: ecology, environmental problems, energy sources, mechanisms for production of electricity, parameters and measurements that can be obtained from field investigations, sociopolitical issues and personal actions. Table 3.4 presents the total count and mean per note of the content codes present in the seven analyzed corpora. The results show some big variations across the discourses in terms of

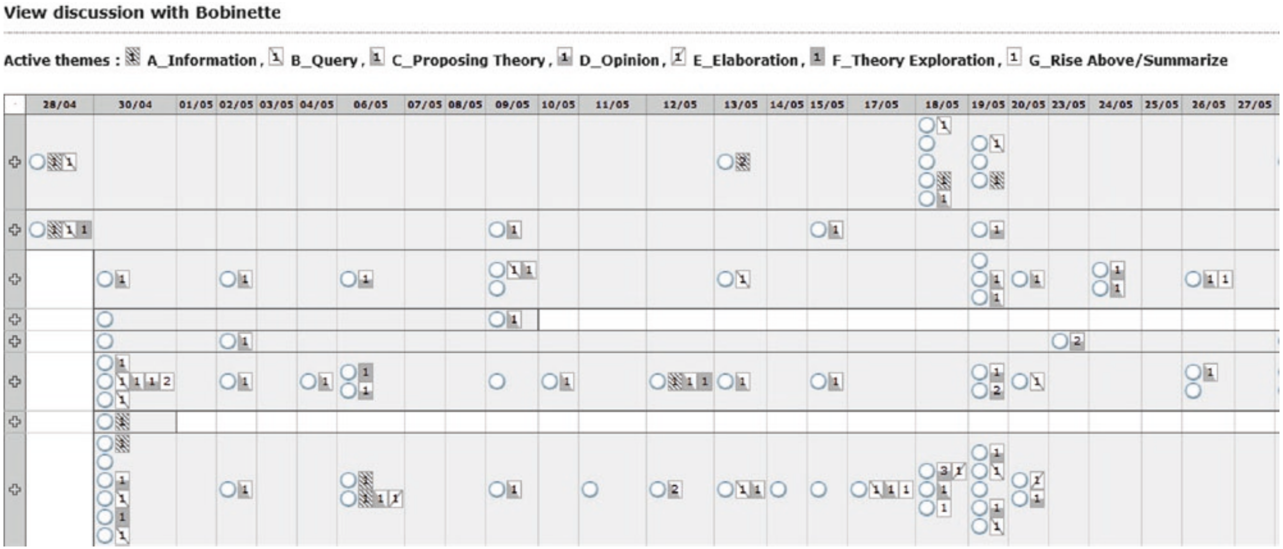

Fig. 3.3 A section of the Bobinette display of the presence of scaffolds in the first 8 threads in the Y78Z9 corpus. Each circle represents one note contributed to the view. Each patterned box following a note represents the presence of that particular type of scaffold in that note, and the number within the box indicates the number of that kind of scaffold within that note

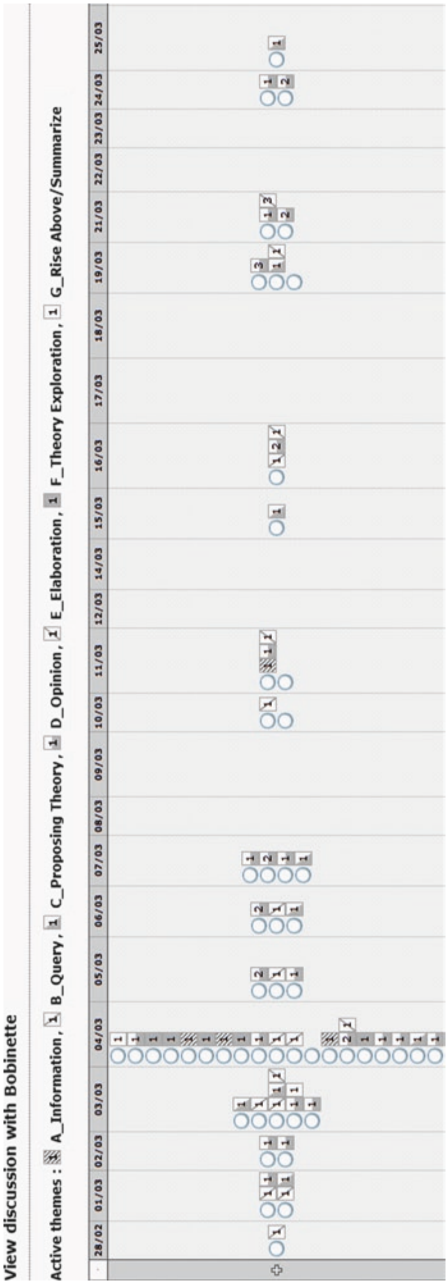

Fig. 3.4 A section of the Bobinette display of the presence of scaffolds in the first thread in the X56QIP corpus. Each circle represents one note contributed to the view. Each patterned box following a note represents the presence of that particular type of scaffold in that note, and the number within the box indicates the number of that kind of scaffold within that note

Table 3.4 The total number and mean number per note of the content codes present in the analyzed corpora

Corpus ID and label	Ecology (mean # per note)	Field investigation (mean # per note)	Environmental problems (mean # per note)	Energy sources (mean # per note)	Mechanisms (mean # per note)	Sociopolitical (mean # per note)	Action (mean # per note)
1. X6A	18	1	29	131	4	18	6
	(0.09)	(0.00)	(0.14)	(0.65)	(0.02)	(0.09)	(0.03)
2. X6B	18	5	89	89	5	28	5
	(0.11)	(0.03)	(0.53)	(0.53)	(0.03)	(0.17)	(0.03)
3. X6C	11	2	20	71	1	18	2
	(0.08)	(0.01)	(0.14)	(0.51)	(0.01)	(0.13)	(0.01)
4. X6_07	50	0	73	312	46	47	9
	(0.24)	(0.00)	(0.34)	(1.47)	(0.22)	(0.22)	(0.04)
5. Y7C	11	0	23	95	9	18	3
	(0.06)	(0.00)	(0.16)	(0.53)	(0.05)	(0.10)	(0.02)
6. Y78Z9	49	11	220	41	8	83	41
	(0.25)	(0.06)	(1.13)	(0.21)	(0.04)	(0.43)	(0.21)
7. X56QIP	1,716	54	356	94	3	31	8
	(1.35)	(0.04)	(0.28)	(0.07)	(0.00)	(0.02)	(0.01)

their content coverage. The first three corpora had a relatively low concentration of content codes. X6_07 and Y7C had a strong focus on energy sources, Y78Z9 had a strong focus on environmental problems, followed by sociopolitical issues while X56QIP was strongly centered on ecology, followed by environmental problems.

We focus our examination of the content codes on the two most productive discourses based on the previous analyses – Y78Z9 and X56QIP – to explore whether there is any structural differences between them in terms of content coverage beyond differences in discussion foci. Figures 3.5 and 3.6 are segments of the Bobinette displays of the content codes for these two respective corpora. In Fig. 3.5, most of the notes have content codes for environmental problems, and many of these also have multiple types of content codes. For example, thread 6 (T6 in Fig. 3.5) started with discussions about environmental problems and ecology, followed by a number of discussions of environmental problems in conjunction with sociopolitical issues and personal action. This seems to indicate that students were very quick to move from the problems to discussions about solutions and there is no difference in the content patterns from the beginning to the end of a thread. On the other hand, Fig. 3.6 shows a very different content topic distribution pattern. The two threads displayed in Fig. 3.6 are typical of the discourse found in X56QIP – most of the notes only had content codes in ecology, many of which had more than one content code in ecology. This seems to indicate that students had a strong focus on exploring ecology topics in depth. When there were other content codes included, most of these were related to environmental problems if these occur earlier in the thread, while the later notes seem to move on to discussions about measurements and parameters from field investigations, different energy sources and a few notes moved on to discuss sociopolitical issues. However, notes with multiple code types are few, and notes with more than two types of content codes are relatively rare, which is in stark contrast to the distribution in Y78Z9 as displayed in Fig. 3.5 (i.e. most of the circles, representing notes, are followed by one colored square, indicating the presence of one type of content codes, and circles followed by more than two colored boxes are rare).

In examining the content of the threads in greater detail, we find that most of the extended threads in Y78Z9 were able to make advances in understanding by drawing on some relevant connections early in the discussion, but cannot make clear advances beyond that point. An illustration of this is taken from some notes in Thread T6. Note A, in discussing how to solve the problem of global warming, proposed "*Can we use some method such as collect excess carbon dioxide and put them in underwater to relax the problems? Although it can't remove the problems, we can have more time to invent new methods before the end of the Earth. http://www.sciscape.org/news_detail.php*". Other students then followed up with questions about cost and whether it would be better to fund scientists to solve the problem instead of just "relaxing" [meaning reducing] the problem. This was followed up by Note B, offering the view that "… *many countries have economic losses due to global warming and these losses are increasing each year. I think it is more worthwhile to use this method than let the scientists slowly invent new solution as invention cost a lot of time. It may even let the government lose more while we are waiting*

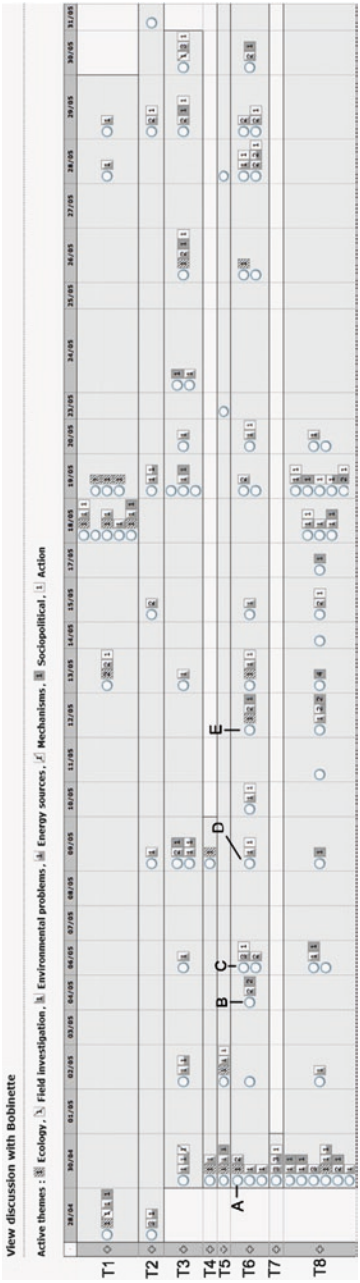

Fig. 3.5 A section of the Bobinette display of the presence of content codes in the first 8 threads (T1 to T8) in the Y78Z9 corpus. Each circle represents one note contributed to the view. Each patterned box following a note represents the presence of that particular type of content code in that note, and the number within the box indicates the number of that kind of content code within that note

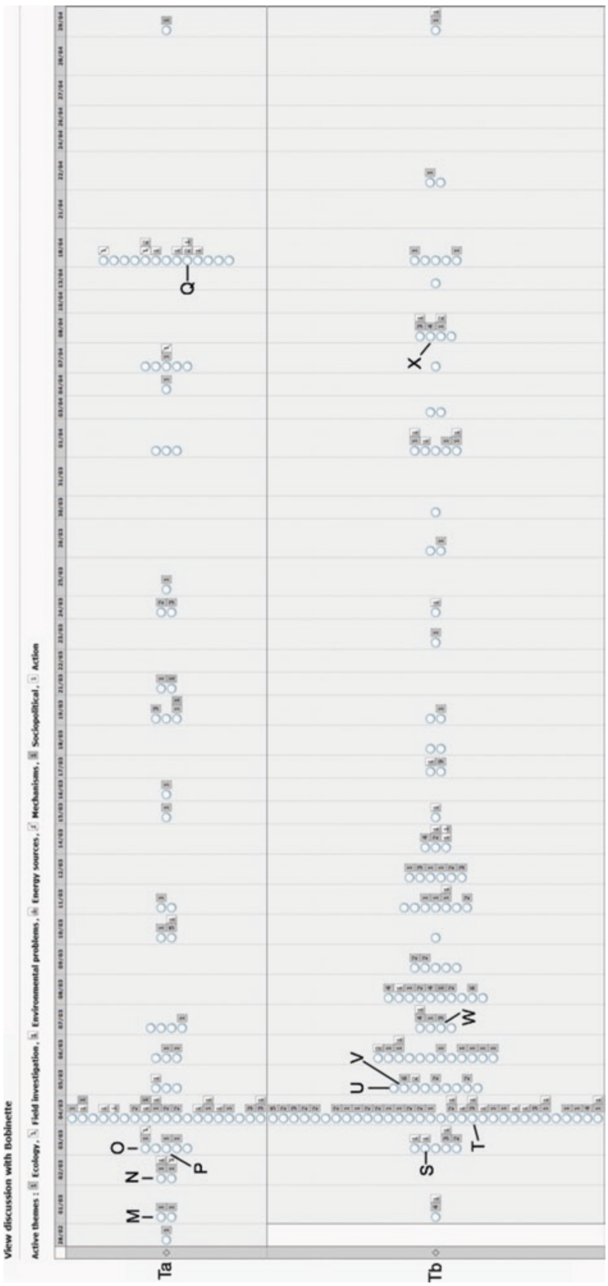

Fig. 3.6 A section of the Bobinette display of the presence of content codes in the first 2 threads (Ta and Tb) from X56QIP corpus. Each circle represents one note contributed to the view. Each patterned box following a note represents the presence of that particular type of content code in that note, and the number within the box indicates the number of that kind of content code within that note

for the scientists. Second, [even] if only those developed countries join this project, the problem of global warming can be greatly relaxed [reduced] since they are the major greenhouse gases producer. These countries are wealthy enough to afford this project." So the discussion moved from proposing a specific method to tackle the problem – collecting carbon dioxide and putting it under water – to talking about costs and government involvement. Note C challenged the feasibility of collecting and putting carbon dioxide under water, and proposed instead taking personal action to reduce greenhouse gas production as the way forward: "… *I want to understand how you can put carbon dioxide underwater. It would come back to the water surface again, right. Secondly, I don't think doing this really solve the basic of the problem. Is it better to reduce the greenhouse gas discharged by us?*" The remainder of the thread was concerned with the need to reduce greenhouse gas production and Note D introduced a sociopolitical dimension by pointing out that "*… We should also try hard to negotiate with America because America give the most carbon dioxide*." The author for Note A came back with more explanations of how carbon dioxide can be trapped and placed under water in liquefied form in Note E. There were no new ideas introduced in the rest of this thread. This pattern of discourse and types of concerns expressed are replicated in the other threads in this corpus.

The discussions in the X56QIP corpus seem to be able to go further in exploring specific problems in greater depth and that different concerns were reflected in the different threads. For example, in thread Ta in Fig. 3.6, the focus was on understanding why some of the organisms recorded as inhabitants in a local stream in Hong Kong sometime ago cannot be found in the stream anymore. Note M asked "*… nowadays HK's water quality was not very good, some of the organism which can be found at the stream at past can't be found nowadays: What is your opinion of Why do the no. of species found in HK stream become lesser?*" Note N responded, identifying pollution as the main culprit: "*I think that HK people keep polluting the environment. Some hikers even leave rubbish in the country parks and rivers. The number of species found in HK stream become less because rubbish will affect the water quality*." Note O then proposed to collect some environmental data to explore the problem further: "*My theory__ we need to collect the pH value in HK hill streams. It is important because different pH value will affect different organism. Also if it is too acid or too basic, it is not a good place to live. We can also know whether the human being affect the hill streams by getting the pH value from the hill streams. We can go to some hill streams in HK to collect the pH value*". Note P extended the scope of the discussion to suggest that global warming and its impact on the food chain may also contribute to the disappearance of certain species, and asked if ecological damages can be reverted by subsequent improvement: "*I need to understand__ I think reason of the number of species found in HK become lesser isn't just only the water quality. It has a lot of reasons like Globe Warming, Food chain have been destroyed etc. What if the water quality become better and better in the future. Will the species which we could see them in the past come back to HK again?*" Discussion continued to explore the ecological impact of various human activities, and bringing in at a later point the environmental impact of factories and using fuel for energy, as exemplified by Note Q: "*My theory is that acid rain is caused by Global Warming and Global Warming is caused by big factories all over the world and the big oil companies*." Even though the first of the two claims made in Note Q was incorrect and the second was substantiated and that none of the subsequent notes in the thread was able to point this out, the thread does exemplify a deepening exploration and genuine efforts to understand the relationship between ecological problems revealed by data and various human activities without jumping to discuss solutions readily gleaned from the popular literature on global warming.

While thread Ta was on understanding the ecological problem of disappearing species, thread Tb began with a more specific concern – how global warming may affect living things on the highest mountain in Hong Kong, after watching a documentary on the problem produced by a local TV station. Note S made a simple statement, "*My theory__ I think Global Warming will affect the animals habitats.*" Note T followed up by exploring how global warming may affect habitats: "*My theory__ The global warming heat their habitats so the creatures that live in cold places need to migrate to other places.*" Note U offered another possibility to migration as a consequence: "*Different opinion__ (I) think if there are some colder habitat where around their old habitat, they will rather go there than migrate*." The same author followed up with Note V to elaborate further: "*If the giant spiny frogs migrate, it will affect the food chain in it[s] own habitat or it[s] new habitat. In it[s] old habitat, some predator which eat the giant spiny frog will extinct. Then the predator which eat the giant spiny frog's predator will extinct too. In the worse case, the whole food chain where the giant spiny frog old habitat will be destroyed. (2) If the giant spiny frog migrate to another place, it might destroy the balance between the other species in that place. It may be eat all the species in that place. It is the best way not to let the giant spiny frog extinct or migrate? How can we not to let the giant spiny frog extinct or migrate?*" Further discussions followed and the discussion moved from understand the impact of global warming on a specific animal to the ecosystem, as exemplified by Notes W ("*I think it may cause a large impact of the food chain because every species rely on the species. It may cause a big problem if one species die. …. The "ecological balance" losses.*") and X ("*When the climate changes, …. the organisms that live in that area need to move to …. the global warming not only affect the habitat, but also affect the ecosystem in that area*.")

Hence, an in-depth reading of the two corpora reveals different patterns and depth of inquiry. The Y78Z9 discourse appears to be more indicative of extensive inquiry while X56QIP has characteristics more indicative of intensive inquiry, as described by Hakkarainen (1998). Such a fine-grained comparison of the depth of inquiry in these two corpora cannot be made on the basis of the automated analysis and visualization output alone, but has to be supplemented and supported by further analysis and interpretation made by human researchers on the basis of careful reading of the discourse corpora. However, the difference in content code patterns between the two corpora signals something interesting for the researchers to follow up on. Whether a discourse that does not show a pattern of sustained focus in one content code is necessarily limited in its depth of inquiry has to be further explored through the analysis of more discourse corpora. What can be concluded from this part of the study is that a visualization of the development of the discourse in terms of content focus provides further insight in the exploration and analysis process.

3.9 Integrating the Different Comparisons: Towards a Methodology for Understanding Learners' Knowledge Building Trajectory

The goal of this study is to develop a methodology, including the identification of appropriate indicators and tools that can be used to provide a quick, first level assessment of the level of knowledge building reflected in an online discourse conducted on Knowledge Forum® using machine analysis and visualization. The key objectives of developing such a methodology are (1) to build an empirically grounded understanding of learners' trajectories of advancement in knowledge building; and (2) to identify pedagogical and facilitation designs that are more conducive to deeper levels of knowledge building by students. The findings reported in this paper indicate that the indicators and tools selected in the present study are promising starting points in the development of such a methodology. Since the patterns and linguistic features of synchronous CSCL discourse such as chat data is different from that in asynchronous settings, it is doubtful that the findings reported here is applicable to those situations. The methodological implications from this study is not confined to discourse conducted on Knowledge Forum®, but should be applicable to the analysis of asynchronous CSCL corpora if the interest is in identifying and/or comparing learners' knowledge building engagement and progression.

First of all, basic participation statistics at discourse and thread levels can be easily generated by tools such as ATK. Such statistics will be useful in discriminating discourses comprising largely of truncated inquiries from other, more engaging and possibly more productive ones.

Visualization of speech acts coded through the use of argumentation and question markers reveals the presence of questions and argumentative moves such as claims, contrasts and reasons that are more indicative of inquiry oriented engagement. The use of process-oriented scaffolds provided by the CSCL platform can also be another important source of information about the understanding and appropriate use or otherwise of such scaffolds by the learners. Our work also indicate that the use of a linear visualization tool to display the time sequence of discourse markers provide additional evidence on whether the markers indicate sustained inquiry over time. If the CSCL platform provide different scaffolds for use at different stages of inquiry such as in the case of Knowledge Forum®, a linear visualization may shed light on whether there is evidence of cycles of progressive inquiry occurring.

We find that fine grained identification of subject matter content within each note using keyword and keyword pattern matching to be valuable in revealing whether the discussion is likely to be an extensive inquiry or an intensive one (Hakkarainen 1998, 2004), with the latter being more in-depth and productive.

Discourses that show a sustained focus on a specific content area, particularly earlier in an inquiry thread, are more likely to be indicative of intensive inquiry.

While we have found the use of automated coding and visualization of discourse corpora to be very helpful in providing insight on the quality of the knowledge building processes of the students in the seven classroom contexts as reflected by the discourse corpora analyzed, the study also demonstrates clearly the need for automated analyses and visualization to be augmented and validated by a review of the actual discourse data by human researchers. One limitation of the present study is the lack of "inter-coder reliability" between machine coding and human coding. Such a reliability check would be very resource intensive because of the variety of coding and the number of corpora involved. However, since the key findings and interpretations are not made purely on the basis of the automatic analyses, we are confident about the claims made about the comparisons of the knowledge building trajectory for the corpora analyzed.

Another limitation of the present study is that the maximum number of thematic codes the Bobinette tool can display at any one time is limited to eight. This hinders the visualization of multiple sets of codes as is possible in CORDTRA. Further, the smallest time unit that Bobinette can differentially display is 1 day and so cannot really display the sequential relationship among notes created on the same day, which becomes problematic when students are given the opportunity to work on line during class contact time, which is often the case. Also, Bobinette cannot display the actual thread structure of the discourse. Hence one direction of our future work is to look for or develop a linear visualizer with the following enhancements: (1) support for user-definable units of time, which can be variable over different periods of the discourse; (2) support display of thread structure; and (3) the number of thematic codes to be displayed is definable by the user, with the color coding for each also user-definable.

In addition to seeking a more powerful linear visualizer, we hope to be able to experiment with applying data-mining algorithms on the coded discourse since it has the potential of augmenting human visualization in discovering regularities and trends (Reimann et al. this volume) in asynchronous knowledge building discourse.

Acknowledgments This research is supported by an earmarked research grant from the Research Grants Council Hong Kong (RGC Ref: 7440/05 H) and funding from the IT Strategic Research Theme at the University of Hong Kong.

References

Andriessen, J., Baker, M., & Suthers, D. (2003). *Arguing to learn: Confronting cognition in computer-supported collaborative learning environment*. Dordrecht: Kluwer Academic.

Andriessen, J. E. B., & Coirier, P. (1999). *Foundations of argumentative text processing*. Amsterdam: Amsterdam University Press.

Bereiter, C. (2002). *Education and mind in the knowledge age*. Mahwah: Lawrence Erlbaum.

Biggs, J. B. (1999). *Teaching for quality learning at university*. Buckingham: Society for Research into Higher Education and Open University Press.

Biggs, J. B., & Collis, K. F. (1982). *Evaluating the quality of learning: The SOLO taxonomy*. New York: Academic.

Burtis, J. (1998). *The analytic toolkit*. The Ontario Institute for Studies in Education, The University of Toronto: Knowledge building research team.

Card, S. K., Mackinlay, J. D., & Shneiderman, B. (1999). *Readings in information visualization: Using vision to think*. San Francisco: Morgan Kaufmann.

Dyke, G., Lund, K., & Girardot, J. (2009). Tatiana: An environment to support the CSCL analysis process. In *Proceedings of the 9th international conference on computer supported collaborative learning* (Vol. 1, pp. 58–67). Rhodes, Greece: International Society of the Learning Sciences.

Erkens, G., & Janssen, J. (2008). Automatic coding of communication in collaboration protocols. *International Journal of Computer-Supported Collaborative Learning, 3*(4), 447–470.

Giguet, E., Lucas, N., Blondel, F., & Bruillard, E. (2009). Share and explore discussion forum objects on the Calico website. In *Community events proceedings of the 9th international conference on computer supported collaborative learning 2009* (pp. 174–176). Poster presented at the computer supported collaborative learning 2009 (CSCL2009), Rhodes, Greece: International Society of the Learning Sciences.

Goodman, B. A., Linton, F. N., Gaimari, R. D., Hitzeman, J. M., Ross, H. J., & Zarrella, G. (2005). Using dialogue features to predict trouble during collaborative learning. *User Modeling User-Adaptated Interaction, 15*, 85–134.

Gweon, G., Jeon, S., Lee, J., Finger, S., & Rosé, C. P. (this volume). A framework for assessment of student project groups on-line and off-line. In S. Puntambekar, G. Erkens, & C. Hmelo-Silver (Eds.), *Analyzing interactions in CSCL: Methods, approaches and issues* (pp. 293–317) Springer.

Hakkarainen, K. (1998). Epistemology of scientific inquiry and computer-supported collaborative learning. Unpublished thesis for the degree of Doctor of Philosophy. University of Toronto.

Hakkarainen, K. (2004). Pursuit of explanation within a computer-supported classroom. *International Journal of Science Education, 26*(8), 979–996.

Hakkarainen, K., & Sintonen, M. (2002). Interrogative model of inquiry and computer-s collaborative learning. *Science & Education, 11*, 25–43.

Heeman, P. A., Byron, D., & Allen, J. F. (1998). Identifying discourse markers in spoken dialog. Paper presented at the AAAI Spring symposium on applying machine learning and discourse processing, Stanford, March.

Hmelo-Silver, C. E., Jordan, R., Liu, L., & Chernobilsky, E. (this volume). Representational tools for understanding complex computer-supported collaborative learning environments. In S. Puntambekar, G. Erkens, & C. Hmelo-Silver (Eds.), *Analyzing interactions in CSCL: Methods, approaches and issues* (pp. 83–106) Springer.

Janssen, J., Erkens, G., & Kanselaar, G. (2007a). Visualization of agreement and discussion processes during computer-supported collaborative learning. *Computers in Human Behavior, 23*, 1105–1125.

Janssen, J., Erkens, G., Kanselaar, G., & Jaspers, J. (2007b). Visualization of participation: Does it contribute to successful computer-supported collaborative learning? *Computers & Education, 49*(4), 1037–1065.

Koschmann, T. (2003). Computer-supported collaborative learning. In J. W. Guthrie (Ed.), *Encyclopedia of education* (2nd ed., pp. 468–471). New York: Macmillan Reference.

Laferrière, T., & Allaire, S. (2005). Scaffolding student teachers' online discourse for knowledge building purposes. *Technology and Teacher Education Annual, 2*, 939.

Lai, M., & Law, N. (2006). Peer scaffolding of knowledge building through collaboration of groups with differential learning experiences. *Journal of Educational Computing Research, 35*(2), 121–142.

Lamon, M., Reeve, R., & Scardamalia, M. (2001). Mapping the growth of deeply principled understandings in a knowledge building community [Electronic Version]. New directions in knowledge building. Retrieved 14 July 2009, from http://ikit.org/lamon/mapping.html.

Landauer, T., Foltz, P., & Laham, D. (1998). An introduction to latent semantic analysis. *Discourse Process, 25*, 259–284.

Law, N. (2005). Assessing learning outcomes in CSCL settings. In T. Koschmann, T.-W. Chan, & D. D. Suthers (Eds.), *Computer supported collaborative learning conference (CSCL)* (pp. 373–377). Taipei, Taiwan: Lawrence Erlbaum Associates.

Law, N., Lu, J., Leng, J., Yuen, J., & Lai, M. (2008). Understanding knowledge building from multiple perspectives. Paper presented in the workshop on common framework for CSCL interaction analysis, International conference for the learning sciences 2008, Utrecht, Netherlands.

Law, N., Lu, J., Leng, J., Yuen, J., & Lai, M. (2009). *Exploring knowledge building discourse using argumentation markers. Paper presented in the workshop on analysis of computer-supported collaborative learning (CSCL) discourse: Advances in theory and analysis tools, CITE research symposium 2009*. Hong Kong: University of Hong Kong.

Law, N., & Wong, E. (2003). Developmental trajectory in knowledge building: An investigation. In B. Wasson, S. Ludvigsen, & U. Hoppe (Eds.), *Designing for change in networked learning environments* (pp. 57–66). Dordrecht: Kluwer Academic Publishers.

Law, N., Yuen, J., Huang, R., Li, Y., & Pan, N. (2007). A learnable content & participation analysis toolkit for assessing CSCL learning outcomes and processes. In C. Chinn, G. Erkens, & S. Puntambekar (Eds.), *Mice, minds & society: Proceedings of the international conference on computer supported collaborative learning 2007* (pp. 408–417). Rutgers University, New Jersey: International Society of the Learning Sciences, Inc.

Lee, E. Y. C., Chan, C. K. K., & van Aalst, J. (2005). Students assessing their own knowledge advances in a computer-supported collaborative environment. In T. Koschmann, D. Suthers, & T.-W. Chan (Eds.), *Proceedings of computer supported collaborative learning 2005: The next 10 years*. NJ: Lawrence Erlbaum Associates.

Lee, E. Y. C., Chan, C. K. K., & van Aalst, J. (2006). Students assessing their own collaborative knowledge building. *International Journal for Computer-Supported Collaborative Learning, 1*, 277–307.

Pennebaker, J. W., Booth, R. J., & Francis, M. E. (2007). *Linguistic inquiry and word count (LIWC)*. Mahwah, NJ: Lawrence Erlbaum Associates.

Popper, K. R. (1972). *Objective knowledge: An evolutionary approach*. Oxford, UK: Clarendon.

Rahikainen, M., Lallimo, J., & Hakkarainen, K. (2001). Progressive inquiry in CSILE environment: Teacher guidance and students' engagement. In P. Dillenbourg, A. Eurelings, & K. Hakkarainen (Eds.), *European perspectives on computer-supported collaborative learning: Proceedings of the first European conference on CSCL* (pp. 520–528). Maastricht, the Netherlands: Maastricht McLuhan Institute.

Reimann, P., Yacef, K., Kay, J. (this volume). Analyzing collaborative interactions with data mining methods for the benefit of learning. In S. Puntambekar, G. Erkens, & C. Hmelo-Silver (Eds.), *Analyzing interactions in CSCL: Methods, approaches and issues* (pp. 161–185) Springer.

Rosé, C. P., Wang, Y. C., Cui, Y., Arguello, J., Stegmann, K., Weinberger, A., et al. (2008). Analyzing collaborative learning processes automatically: Exploiting the advances of computational linguistics in computer-supported collaborative learning. *International Journal of Computer Supported Collaborative Learning, 3*(3), 237–271.

Saab, N., van Joolingen, W. R., & van Hout-Wolters, B. H. A. M. (2005). Communication in collaborative discovery learning. *The British Journal of Educational Psychology, 75*(4), 603–621.

Salton, G., Wong, A., & Yang, C. S. (1975). A vector space model for automatic indexing. *Communcations of the ACM, 18*(11), 613–620.

Scardamalia, M. (2002). Collective cognitive responsibility for the advancement of knowledge. In B. Smith (Ed.), *Liberal education in a knowledge society* (pp. 67–98). Chicago, IL: Open Court.

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

Scardamalia, M., & Bereiter, C. (1992). Text-based and knowledge-based questioning by children. *Cognition and Instruction, 9*(3), 177–199.

Scardamalia, M., & Bereiter, C. (1994). Computer support for knowledge-building communities. *The Journal of the Learning Sciences, 3*, 265–283.
Scardamalia, M., & Bereiter, C. (2003). Knowledge building environments: Extending the limits of the possible in education and knowledge work. In A. DiStefano, K. E. Rudestam, & R. Silverman (Eds.), *Encyclopedia of distributed learning* (pp. 269–272). Thousand Oaks, CA: Sage Publications.
Scardamalia, M., & Bereiter, C. (2006). Knowledge building: Theory, pedagogy, and technology. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 97–115). New York, NY: Cambridge University Press.
Scardamalia, M., Bereiter, C., & Lamon, M. (1994). The CSILE project: Trying to bring the classroom in to world 3. In K. McGilly (Ed.), *Classroom lessons: Integration cognitive theory and classroom practice* (pp. 201–228). Cambridge: MIT Press.
Scott, J. (2000). *Social network analysis: A handbook*. London: Sage.
Soller, A. (2004). Computational modeling and analysis of knowledge sharing in collaborative distance learning. *User Modeling and User-Adapted Interaction, 14*, 351–381.
Stolcke, A., Coccaro, N., Bates, R., Taylor, P., Van Ess-Dykema, C., Ries, K., et al. (2000). Dialogue act modeling for automatic tagging and recognition of conversational speech. *Computational Linguistics, 26*, 339–373.
Suthers, D., & Medina, R. (this volume). *Tracing interaction in distributed collaborative learning*. In S. Puntambekar, G. Erkens, & C. Hmelo-Silver (Eds.), *Analyzing interactions in CSCL: Methods, approaches and issues* (pp. 341–366) Springer.
Tan, S. C., Yeo, A. C. J., & Lim, W. Y. (2005). Changing epistemology of science learning through inquiry with computer-supported collaborative learning. *Journal of Computers in Mathematics and Science Teaching, 24*(4), 367.
Teplovs, C. (2008). The knowledge space visualizer: A tool for visualizing online discourse. Paper presented at the common framework for CSCL interaction analysis workshop at the international conference of the learning sciences 2008. Utrecht, NL.
Teplovs, C., & Scardamalia, M. (2007). Visualizations for knowledge building. Paper presented at IKIT knowledge building summer institute 2007, Toronto, Canada.
van Aalst, Chan, K. K. C., & Lee, E. (2002). Developing knowledge-building inquiry through knowledge building portfolios in a Hong Kong classroom. Paper presented in the symposium "Knowledge Building: Multicultural Perspectives" at the annual meeting of the American educational research associations, New Orleans.
van Boxtel, C., van der Linden, J., & Kanselaar, G. (2000). Collaborative learning tasks and the elaboration of conceptual knowledge. *Learning Instruction, 10*(4), 311–330.
Veldhuis-Diermanse, A. E. (2002). CSCLearning? Participation, learning activities and knowledge construction in computer-supported collaborative learning in higher education. Unpublished PhD Thesis, Wageningen University, Wageningen.
Zhang, J., Scardamalia, M., Lamon, M., Messina, R., & Reeve, R. (2007). Socio-cognitive dynamics of knowledge building in the work of 9- and 10-year-olds. *Educational Technology Research and Development, 55*, 117–145.

Chapter 4
Representational Tools for Understanding Complex Computer-Supported Collaborative Learning Environments*

Cindy E. Hmelo-Silver, Rebecca Jordan, Lei Liu, and Ellina Chernobilsky

Abstract To learn and reason effectively about complex phenomena, a central goal in education, learners need opportunities to engage with them. Technology-mediated learning environment, such as simulations, can provide opportunities for learners to formulate, test, refine, and repair their mental models of complex systems. However, technology tools do not stand alone – they are situated in complex learning environments that require consideration of learning at the level of the individual, small group and whole class, as well as consideration of the roles that both the teacher and technology play in scaffolding student learning. While technology can enable the development of rich learning environments, the extent to which and how this technology can be used by teachers and how it influences the nature of student's collaborative knowledge construction is still unclear. Appropriate analytical tools are needed to represent students' mediated interactions in computer-supported collaborative learning (CSCL) environments. In this chapter, we present analytical tools that help us construct a comprehensive picture of how learning is mediated over time through a complex interplay of tool use, teacher scaffolding, and collaborative discourse to investigate the mediating roles of technology and teacher and peer scaffolding in CSCL. In particular, we will show how different representations (e.g., CORDTRA diagrams) can provide insight into the complexity of understanding of complex CSCL learning environments.

Computer-supported collaborative learning (CSCL) environments are complex and often require integrating across multiple coding schemes and different sources of data. Such data may be represented as frequencies of discourse acts, content

* This research was funded by an NSF CAREER grant # 0133533 to the first author, NSF ROLE grant # 0107032, and NSF ALT Grant # 0632546. Conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the National Science Foundation.

C.E. Hmelo-Silver (✉)
Graduate School of Education, Rutgers University, 10 Seminary Place, New Brunswick, NJ, USA
e-mail: chmelo@rci.rutgers.edu

S. Puntambekar et al. (eds.), *Analyzing Interactions in CSCL*, Computer-Supported Collaborative Learning Series, DOI 10.1007/978-1-4419-7710-6_4,

references, or logs of tool use. This kind of representation, however, can hide the relation between different discourse acts as well as obscure sequential information. In this chapter, we describe a technique for visually representing multiple sources of data over time and describe how we have used this in both synchronous and asynchronous CSCL settings.

This work is situated in a social constructivist framework which posits that knowledge is socially constructed (Palincsar 1998) and is mediated by tools (Cole and Engeström 1993). This perspective argues that to understand learning in context, it is critical to understand how tools and social activity are part and parcel of the learning process. In CSCL settings, it is particularly important to understand how learners develop shared understanding through interactions with each other and with computer tools. This understanding necessitates examining how discourse unfolds over time and how computer tools mediate discourse.

Both qualitative and quantitative approaches have been used to analyze CSCL discourse. First, analyzing the complexity of collaborations is often conducted using intricate coding schemes that quantify different types of utterances (e.g., Chinn and Anderson 2000; Engle and Conant 2002; Hmelo-Silver 2003). Particularly in CSCL research, content analysis approaches have been used widely to characterize group discussions by coding and counting the frequencies of different aspects of discourse (e.g., Hmelo et al. 1998; Kumpulainen and Mutanen 1999; De Wever et al. 2006; Fischer et al. 2002; Lai and Law 2006; Lajoie et al. 2006; Strijbos et al. 2006). Other quantitative approaches to analyzing CSCL data include experimental methods (e.g., Rummel and Spada 2005; Rummel et al. 2011; Suthers and Hundhausen 2003), multilevel analysis (Cress 2008; Janssen et al. 2007), and social network analysis (e.g., Nurmela et al. 1999).

In contrast to the quantitative approaches, other researchers have used more qualitative approaches to studying collaboration, such as Roschelle's (1996) study of computer-mediated convergent collaborative conceptual change or Stahl's (2006) analysis of the use of a graphical reference tool to support group cognition in an online chat. These qualitative approaches provide a detailed look at a small portion of collaborative learning activity, focusing on social and linguistic processes. The coding-and-counting extended periods of time but lose sequential detail and often, the relation between different kinds of discourse moves and how those relate to the tools in the learning environment.

Because we, however, are interested in how the different aspects of discourse relate to each other over time in addition to the tools being used in the collaborative learning process, we need to go beyond the coding and counting frequencies of individual speech acts to a consideration of longer sequences of speech. This necessitates that researchers consider the chronological dimension of learning (Chiu 2008; Jeong et al. 2011; Mercer 2008; Reimann 2007) that are often missing in analytic approaches based on aggregated frequencies.

We view the learning environment as an activity system (Engeström 1999) where learners engage and interact with their environment to transform particular objects of activity to achieve an outcome, which is mediated by cognitive and physical artifacts. In understanding this system, we need to investigate the cultural tools that are used – the language used in discourse and the computer technology that embodies particular pedagogical models. Because the nature of group interactions is ultimately a dynamic process (Arrow et al. 2005; McGrath et al. 2000), understanding time is also critical in the analysis of collaborative learning (Reimann 2007). To help us make sense of the complex multidimensional data associated with multiple elements of discourse over time, we have found that visual representations can be important tools for studying CSCL.

4.1 Visual Representations for Understanding CSCL

Researchers have incorporated time into their analyses in several ways. De Laat and Lally (2003) examined how discourse changed over time in an online discussion using a combination of critical event recall and content analysis over different discrete phases of an online discussion. Zemel et al. (2007) developed an emergent coding scheme based on critical events as they unfolded during sequential online chats. Using multivariate statistics and chronologically-oriented long sequences to visualize the data, these individuals were able to draw conclusions about group interactions. Schümmer et al. (2005) combined discourse with log data as they addressed the difficulty of representing these data for analysis. Another approach to examining how CSCL unfolds over time is through uptake analysis diagrams, which provide a visual representation of how ideas are taken up in small group discourse (Suthers et al. 2009; Suthers and Medina this volume); Luckin (2003); Luckin et al. (2001) presented a means to visualize their data through CORDFU (Chronologically-Ordered Representation of Discourse and Features Used). They used this approach to examine how alternative ways of structuring hypermedia affected collaborative discourse, which allowed them to explore relations between the software's features and collaborative knowledge construction in a chronological sequence.

This short review shows that analyzing the complexity of CSCL environments is a difficult undertaking that often requires integrating across multiple data sources. A diagram may be much easier to interpret than presentations of the same information in a verbal form. Visual representations can aid in interpreting complex patterns. Therefore, it is not surprising that other researchers have looked for visual representations to help in understanding the complexity of new learning environments. For example, in a mixed-methods analysis of CSCL interactions using quantitative measures, social network analysis, and qualitative data, researchers provided a visual representation of interaction patterns (Martinez et al. 2003). To capture collaboration in a face-to-face discussion, Strom et al. (2001)used directed graphs to map the semantic space of instructional discourse as students coordinated conceptual and procedural knowledge. In a small group CSCL environment, Suthers et al. (2009; this volume) developed uptake graphs to examine how ideas flowed during collaborative knowledge construction as pairs of students engaged in online discussions. Other researchers have developed analytic tools that focus on how a concrete artifact is constructed. In particular, Avouris et al. (2003) developed the OCAF (Object-oriented Collaboration Analysis Framework) to work backward from a collaboratively created diagrammatic artifact and study the history that led to the construction of that artifact. This framework considers the artifact as a solution to a problem that is constructed by actors who have ownership of different entities that contribute to the development of the artifact. Although each of these methods illuminated different aspects of the collaborative learning processes in CSCL environments, the methods were not completely transparent in terms of constructing the representation. In addition, these methods were not necessarily developed for studying learning over extended time periods. In the remainder of this chapter, we will explain how the CORDTRA diagrams were constructed and demonstrate how it was used in two different CSCL settings.

Chronologically-Ordered Representation of Discourse and Tool-Related Activity (CORDTRA) is a generalization of CORDFU, which enables us to combine the chronological picture of the coded discourse with other learning activities that might be represented as log data or other kind of information coded from video (e.g., drawing activity, gestures). We argue that CORDTRA can foster holistic visualization of the data while enabling fine grain coding. We describe CORDTRA in detail below, but we will first address various approaches to analyzing collaborative computer supported learning environments (CSCL) for which CORDTRA analysis is well suited.

4.2 CORDTRA Diagrams

CORDTRA diagrams contain a timeline where multiple processes are plotted on one timeline in parallel. This allows a researcher to juxtapose a variety of codes to understand an activity system—for example, these might be discourse, gestural, or tool-related codes as shown in Fig. 4.1. Initially, we used the CORDTRA technique to examine face-to-face collaboration in a problem-based learning (PBL) tutorial to understand how constructing a drawing mediated learning (Hmelo-Silver 2003). In this work, a multidimensional coding scheme was used to code discourse at a fine grain of analysis to capture different features of the discourse. These discourse features were initially used to compile frequency counts and later, used as part of the CORDTRA diagram. In addition, these features were used along with information on drawing activity and gestures. Together these analyses allowed us to get a sense of how the PBL tutorial unfolded and what role an external representation played in mediating collaborative learning. We have since used this technique to represent collaborative processes in asynchronous CSCL contexts (Hmelo-Silver et al. 2008, 2009) as well as in synchronous contexts (Hmelo-Silver, Liu, and Jordan 2009). The CORDTRA diagrams were often used to provide the basis for contrasting case analyses, as in the two examples that we present in this chapter.

The CORDTRA diagrams are constructed by creating a unified transcript that integrates the log file data of all the tool hits with the coded discourse data. These discourse data are recorded as number of turns as shown in Fig. 4.1. Technically,

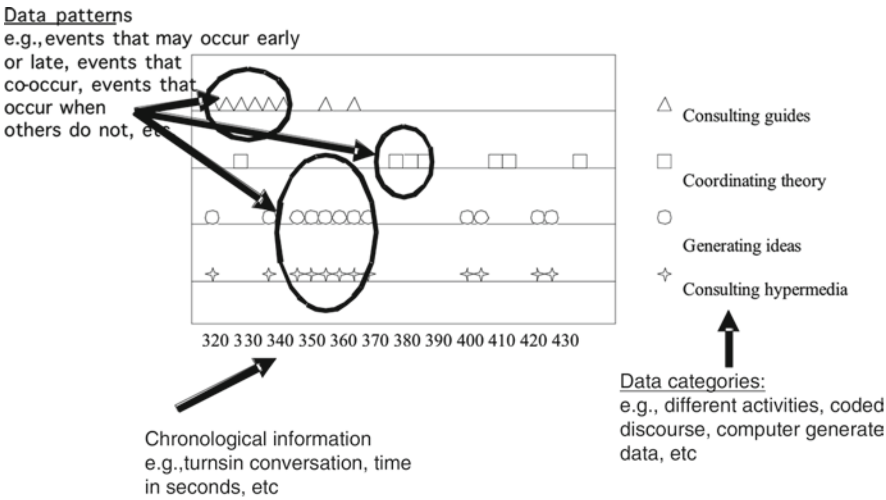

Fig. 4.1 Explanation of CORDTRA (Hmelo-Silver et al. 2008)

these diagrams are scatter plots created in a commercial spreadsheet program. These diagrams can include as many or as few coding categories and sequential logs of tools as needed to study a particular question (Chernobilsky et al. 2003, April). It is up to the analyst to decide what information will be displayed in the diagram as well as for what range of events. Studying the CORDTRA diagrams often suggests points in the discourse that are in need of further investigation. In conducting these analyses, researchers are able to cycle back and forth between the CORDTRA diagrams and actual discourse data.

4.3 Example I: CORDTRA Analysis for Asynchronous CSCL

The context for this analysis was a class in Educational Psychology. The learning goal for the course was to help students develop an understanding of how educational psychology principles applied to classroom instruction. The course was run as a problem-based class supported by the STELLAR system (see Derry et al. 2006 for details). This involved preservice teachers as the students in a hybrid course with both face-to-face and online interactions. For this analysis, we compared and contrasted two groups: one that collaborated extremely well and another that was less successful in their collaboration. We defined Group 1 as being less effective collaborators based on students' self reports of what they learned from the interaction at the end of each problem. In addition, although the students made an effort at engaging with the STELLAR system, they required a great deal of scaffolding, as we will discuss later. We defined group 2 as being more effective. Group 2 was studied in the past because they worked very well together (Hmelo-Silver, Katic, Nagarajan, and Chernobilsky 2007a). Their ratings of the interactions started out comparable to Group 1 after the first problem but they continued to rise until they had the highest ratings in the class for overall interaction after the third problem.

4.3.1 Context for Study: The STELLAR Learning Environment

STELLAR (Socio-Technical Environment for Learning and Learning Activity Research) is an integrated online PBL environment for preservice teachers (Derry et al. 2006) that includes a learning sciences hypermedia (the Knowledge Web), a library of videocases, and an online activity structure that gives students and instructors access to a suite of individual and collaborative tools. The individual tools include a personal notebook that the students use to (a) conduct preliminary problem analysis, (b) keep notes on their research, (c) provide an explanation of their group's product, and (d) reflect on their learning experience. The collaborative tools include the STELLAR whiteboard and threaded discussion. The whiteboard served as the editable solution space, where students could post and edit their solution proposals during and after discussions. This is where students outlined the text for their problem solutions and where other students and the facilitator could comment on proposals and ask questions.

The students engaged in three online problem-based activities. For each activity, they redesigned all or part of a lesson presented in a video case. In this chapter, we focus on the collaborative processes during the second online problem. We focused on the second online problem because the students were familiar with the STELLAR tools by that time. This problem required that students use several online resources as they viewed two contrasting video cases. One case showed a traditional physics teacher who used lectures and demonstrations. The other video case, showed a constructivist instructional approach. This second video served as a contrasting case to help students differentiate instructional approaches (Derry 2006; Schwartz and Bransford 1998). The students were asked to help the first teacher adapt some of the techniques the second teacher used in order to improve a lesson on static electricity. The group used the eight-step activity structure as described in Derry et al. (2006). This structure involved individual and collaborative phases. The students spent 3 weeks on this problem.

4.3.2 CORDTRA Analysis

The conversations that occurred during the 3-week unit were coded for content, collaboration, complexity, questioning, justification, and monitoring. These categories were chosen because they serve as indicators of cognitive engagement (Hmelo-Silver 2003). Each category was further broken into subcategories as discussed in Hmelo-Silver et al. (2008). During the activity, students used several

STELLAR tools, including the Knowledge Web and research library, video cases, whiteboards and discussion boards, online help, personal notebooks and the screen that allowed the students to view other students' initial solutions to a problem. We also examined the STELLAR system log data, which contains records of all the tools accessed and all entries made by the students and facilitators in their personal notebooks, threaded discussion, and group whiteboard. We did so to investigate the discursive contexts in which the tools were used as we constructed the CORDTRA diagrams.

Although we have found that frequencies provide one view of what is going on in the groups, they neither provide a sense of chronology nor inform the researchers about particular qualitative features of the discourse (see Hmelo-Silver et al. 2008 for an example). In the CORDTRA diagrams shown in Figs. 4.2 and 4.3, the data are arranged in chronological order on the horizontal axis. At the bottom of each diagram, there is a running count of lines of codes. For ease of presentation, we show only the collaborative phases of activity for both groups. Since we are choosing to analyze only steps 3 – 6 of the activity for each group, the line counts begin at different numbers for each group because Group 2 engaged more with the STELLAR tools in the earlier phases of the activity than did Group 1.

The vertical axis shows the categories of tool hits, discourse codes, and speakers. The horizontal axis shows the number of tool-related events, either a log entry or a discourse turn. The bottom seven categories represent tool hits (i.e., log data) by any member of the group. The top six or seven categories represent the speakers. The remaining categories represent discourse codes.

One point distinguishing the two groups from the outset is how the groups engaged the facilitator (a TA) and the course instructor (CHS, who also helped facilitate). In contrast, in Group 1, the facilitators were involved early and fairly frequently, and asked most of the explanatory and metacognitive questions. Ann and Fauna seemed to dominate the discourse though other students contributed. In Group 2, the facilitators joined in much later and made infrequent contributions in the form of questions. This is because in this group the students themselves asked a number of questions throughout the duration of the problem. The group participated fairly evenly except for Matt who joined in late in the group's work. His contribution built on one of the other student's ideas and was grounded in personal experience. His later contributions offered an important new idea for an activity that was grounded in psychological theory.

Another aspect of how the CORDTRA diagrams help us distinguish the collaborative activity between two groups is by showing the overall relation between the discourse and the tool use. In Group 1, the students initially viewed the video and the Knowledge Web but after about line 650, none of the group members used these two resources until the very end of the collaborative phase at line 1,000. The content of their online postings were intermixed with conceptual, social, task and tool-related talk throughout their work on the problem. There was some discussion of tools as a problem midway through the discussion. In contrast, Group 2 engaged in very little social and tool related talk—most of their discussion was conceptual with a small amount of task talk sprinkled throughout. What is interesting in Group 2

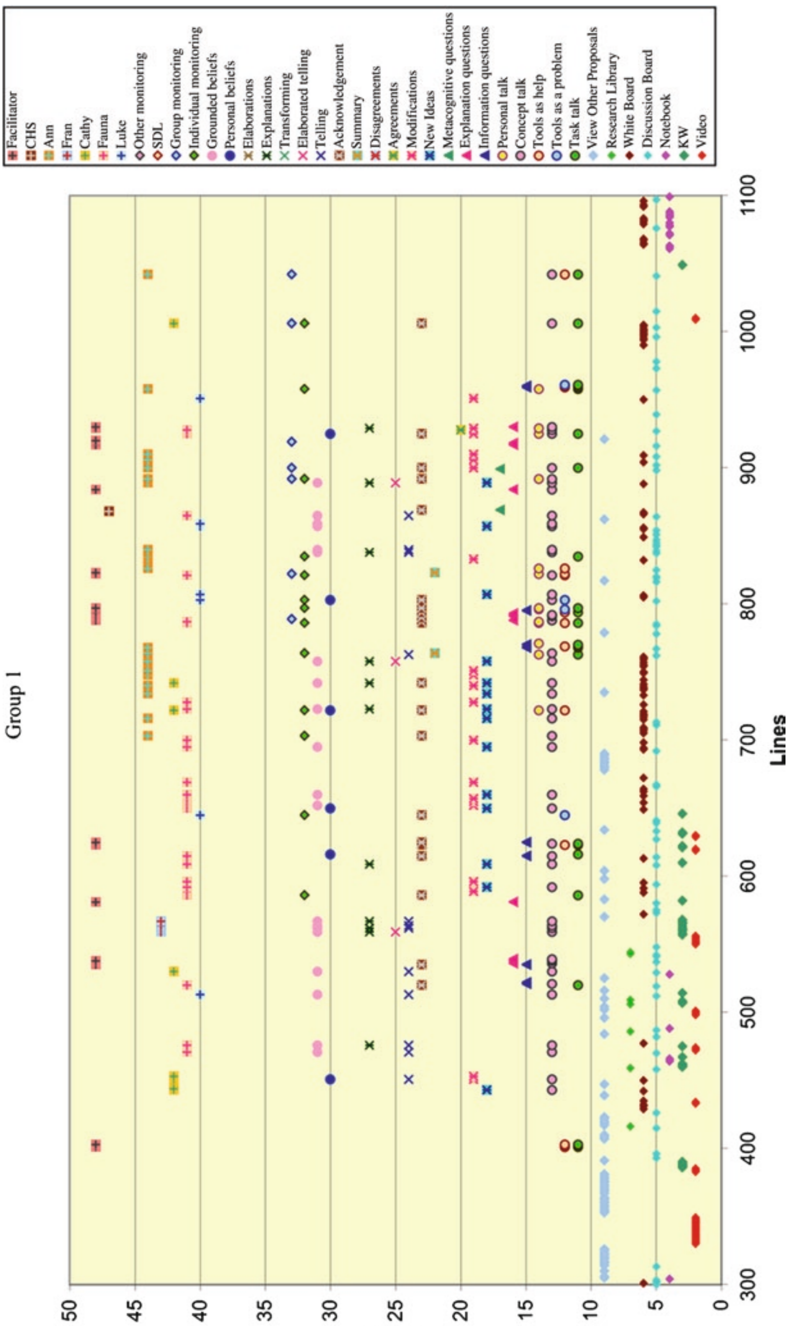

Fig. 4.2 CORDTRA Group 1 (Hmelo-Silver et al. 2008)

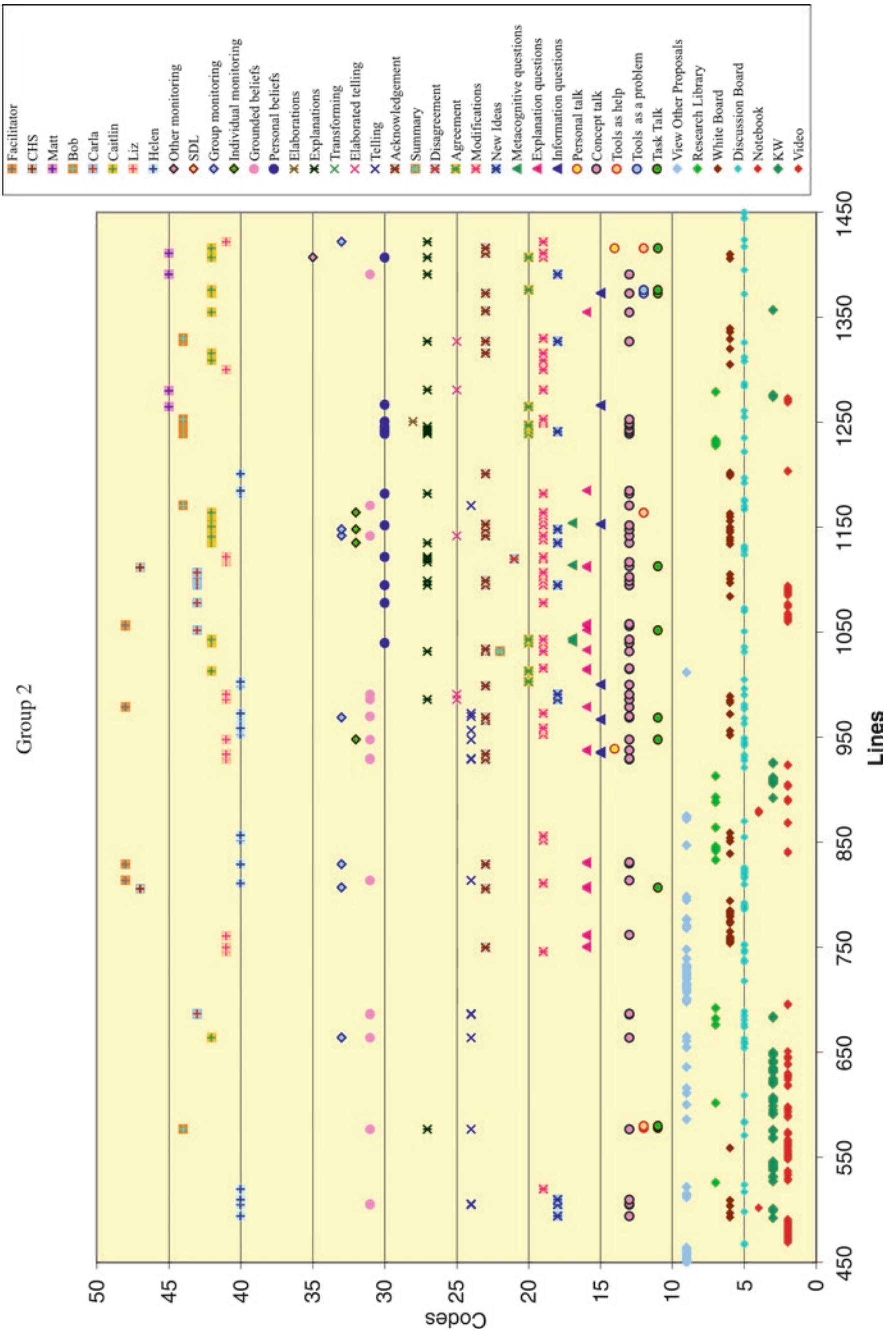

Fig. 4.3 CORDTRA Group 2 (Hmelo-Silver et al. 2008)

is that they went back to the video and the Knowledge Web at intervals throughout the discussion (e.g., around lines 850–925, 1,050–1,100, 1,275). It appears that group members did this following several explanation questions as Fig. 4.3 demonstrates (for example at about line 1,050). This suggests that Group 2 was using the resources of the video and the Knowledge Web in a purposeful manner. That is, they seemed to be using the STELLAR resources to answer the questions at hand. They were bringing together the problems of practice and conceptual ideas repeatedly, which we hypothesize is a precondition for transfer (Derry 2006).

The CORDTRA demonstrates several other distinctions between the two groups with regards to timing. Group 2 engaged in group monitoring throughout the activity whereas Group 1 did not engage in this kind of monitoring until relatively late in the activity. This was generally a request for feedback from the rest of the group. Inspection of the data shows that it was the facilitator doing most of the group monitoring in Group 1 whereas in Group 2, it was the students who were the ones monitoring themselves. The facilitators adapted their support to the needs of the group and thus differentially engaged with each group as needed.

CORDTRA diagrams also allow zooming in and out on different parts of the activity. Inspection of the diagrams displayed in Figs. 4.4 and 4.5 suggest that there were at least two phases of activity for each group. For Group 1, zooming in to lines 400–600 (Fig. 4.3) makes it clear that the students were looking at the discussion and whiteboards but posting little. Most of their posting in this early phase was in the discussion boards where they were “dumping” their research with little processing. The posts coded as “telling” indicate this. For the sake of space, we do not include a CORDTRA of the second phase but that diagram shows that the students were generating few new ideas and many of the modifications were of individual student’s own ideas in response to the facilitator.

Zooming in on Group 2, as shown in Fig. 4.5, there appear to be three phases of activity that correspond to the three middle parts of the activity structure. The first phase was from approximately lines 450–740. This view shows that these students posted some new ideas in the whiteboard. Inspection of the posts shows that these were early proposals for assessments and activities. As they were working on these, there was some lurking, particularly by Matt, but in between the posts, the students used the Knowledge Web and video. The level of discussion in this phase was low. There were some new ideas and some knowledge telling in the discussion board. For the class as a whole, that was characteristic of the discussion board. In this phase, the group was involved in generating some tentative ideas and researching the concepts they were exploring. The group seemed to move into the second phase when Liz asked some explanation questions. For example, in response to one of the whiteboard posts for an assessment (i.e., evidence of understanding) of discussions she commented, “how will you structure the discussion? How is this an evidence of understanding?” In this section of early development of an instructional plan, there were many proposals and questions about the proposals. Most of these were by the students. The students went back and forth between STELLAR resources and their questions.

The third phase of final proposal development began in earnest at about line 930, shown in Fig. 4.5. In this final phase, all the students were actively engaged. They

Fig. 4.4 Beginning in Group 1 (Hmelo-Silver et al. 2008)

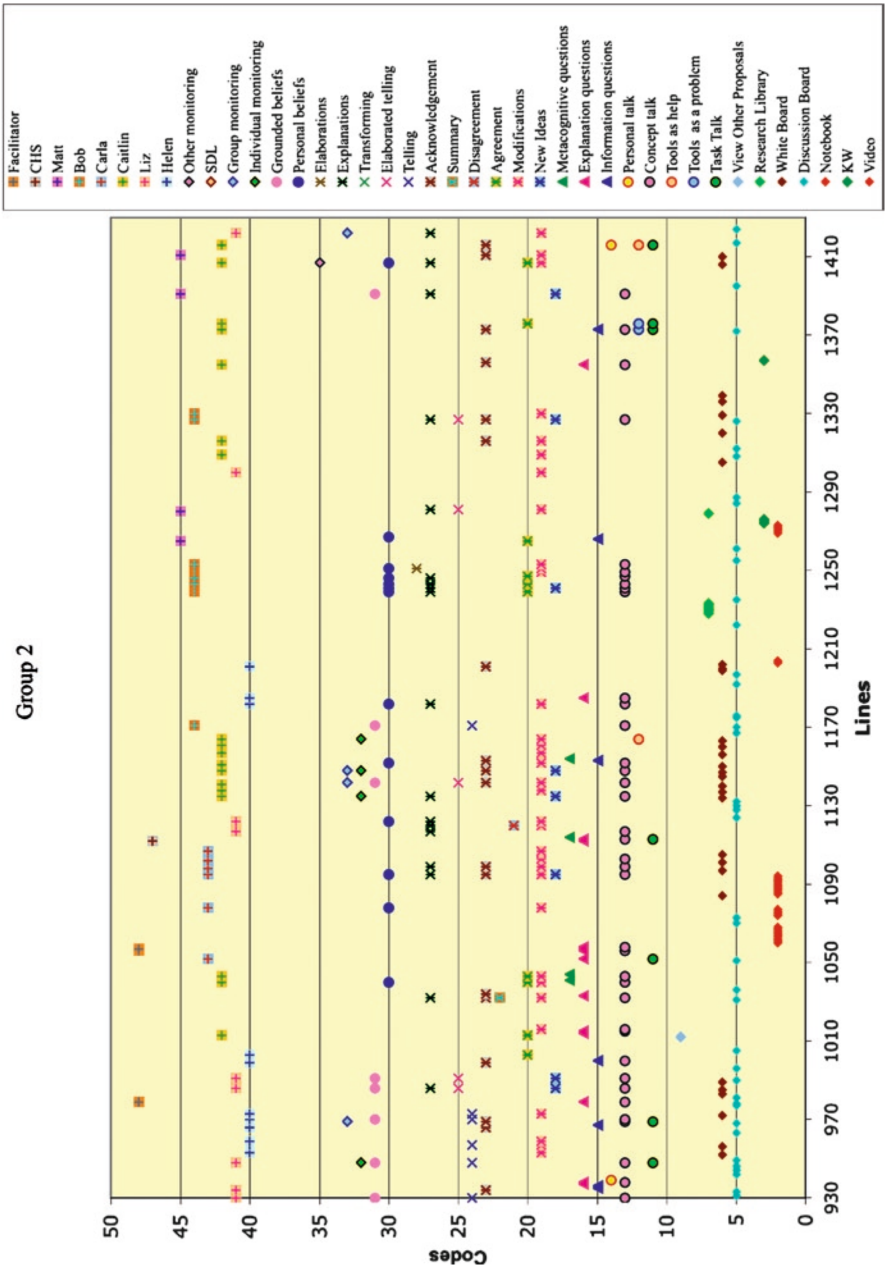

Fig. 4.5 Group 2 Finishing up (Hmelo-Silver et al. 2008)

were focused on building the explanations that would be part of their instructional plans. The group members were going back and forth between the whiteboard and discussion board. Inspection of the online discourse indicates they tried to make connections between their research and plans as in this comment from Matt who made a suggestion about using a computer simulation as part of the physics lesson:

> Introducing some type of software modeling tool on a computer may be also helpful activity the software based activity allows for a variety of testing in different [sic] situations. The program would also make it easier for the students to compare their results from previous [sic] tests. It would be more practical to conduct the activity occasionally, possibly one day out of the week for the lesson. This type of activity is useful because the student are also able to see where they are in the learning process and are able to reflect on what there own inquiries [sic] and the other students inquiries. this type of tool was very useful in one study where students from grades 7–9 were able to outperform high school students from grades 11–12 when it came to applying basic principles to real-world situations(how people learn 217) [sic]. This type of program allows for students to be introduced to a wider range of testing then they would be introduced to in a regular class.

Here Matt made a suggestion he justified based on research he had read. This suggests that although Matt appeared to be lurking during much of the discussion, he was paying attention and posted at an important point in the discussion. In general, the group worked on the whiteboard until someone asked a hard question and then they worked out issues in the discussion board until they were ready to finish up. The CORDTRA helped us realize that Matt's contribution was a change in his online behavior and suggested it was worth additional analytic attention.

In sum, examining CORDTRA diagrams allows us to examine the activity system from various angles: Student collaboration during the discourse, the relation between the discourse and the tools students used while collaborating, and the facilitator input during the activity. Studying multiple angles simultaneously allows us to glean some insight on how more and less effective groups approach collaboration in complex asynchronous CSCL environments and make some conclusions outlined in the discussion section of this chapter.

4.4 Example II: CORDTRA for Synchronous CSCL

The second example is drawn from students working synchronously around a computer simulation in the context of the RepTools suite (Hmelo-Silver et al. 2009). In the first example, we used the CORDTRA technique to juxtapose log data and coded discourse; in contrast, this example focuses on looking at the relationship between different kinds of discourse moves and the content that the students discuss as they work with a computer tool. The RepTools suite includes computer-based representational tools for inquiry into complex biological systems. This example focuses on aquaria as a model for a closed ecosystem. The design of our instructional intervention was informed by structure-behavior-function (SBF) theory, which originated in artificial intelligence research (Goel et al. 1996). For example, in an aquarium, structures refer to components of a system (e.g., fish, plants, filter, water).

Behaviors refer to how the structures of a system achieve their output. These are the interactions or mechanisms that yield a product, reaction, or outcome (e.g., bacteria remove waste by converting ammonia into harmless chemicals). Finally, functions refer to the role or output of an element of the system (e.g., lights provide energy). SBF theory suggests that by considering structures, behaviors, and functions, one can reason effectively about complex systems, and indeed, in the aquarium domain we have demonstrated that experts reason in ways consistent with SBF theory (Hmelo-Silver, Marathe, and Liu 2007c).

The RepTools toolkit includes a function-oriented hypermedia and two NetLogo computer simulation models (Wilensky and Reisman 2006). We use two NetLogo simulations – the fish spawn model and the nitrification process model – to provide models of aquaria at different scales. The fish spawn model at a macrolevel, simulates population density because of fish spawning within an aquarium. The purpose of the model is to help students learn about the relationships among different aspects of an aquarium ecosystem, such as the amount of food, initial gender ratio, filtration, water quality, reproduction, and fish population dynamics. The nitrification process simulation presents a microlevel model of how chemicals reach a balance to maintain a healthy aquarium. This simulation allows students to examine the bacterial-chemical interactions that are critical for maintaining a healthy aquarium. In both NetLogo simulations, students can adjust the values of variables such as fish, plants, and food and observe the results of the adjustment, by sliders. Counters and graphs provide alternative representations for students to examine the results of their inquiry. Alongside a physical aquarium installed in the classroom, students could observe the simulations, generate hypotheses, test them by running the simulation and modify their ideas based on observed results.

As students work with the RepTools materials, they have demonstrated substantial learning gains on pre and post tests (Hmelo-Silver et al. 2007b; Liu 2008). Although those results demonstrated that students did indeed learn about aquaria, we wanted to better understand how they learned by examining the relationship between collaborative learning processes, epistemic practices, and content understanding. We will present examples of two groups drawn from a larger study by Liu (2008). The participants in the larger study were middle school students from two public schools who volunteered to participate in this study. Seventy were seventh graders taught by Ms. W. Seventy five were eighth graders taught by Mr. K. These students were randomly assigned into groups by their teachers and 20 focal groups' interactions were video and audiotaped. The study was conducted in seventh and eighth grades as part of students' science instruction. In this analysis, we contrast two of these focal groups, a high achieving group from Mr. K's class (group 1) and a low achieving group from Ms. W's class (group 2). We selected these groups because although they were both engaged and began with similar pretest scores, posttest scores indicated that one group (group 1) demonstrated considerable conceptual growth and the other (group 2) did not (see Liu 2008 for details).

Both teachers used the unit for approximately 2 weeks and succeeded in getting students engaged in most of the learning events. In both classrooms, before using the computer simulations, both teachers started with a class discussion on the aquarium ecosystem to activate students' prior knowledge and make connections to the physical aquarium in the classrooms. Then the teachers introduced the hypermedia. Students explored the hypermedia software in groups followed by other activities such as class discussions and construction of concept maps that connected parts of the system to their function. The students then collaboratively explored the fish spawn simulation and the nitrification process simulation as they tried to construct a deeper understanding of the aquarium ecosystem.

4.4.1 CORDTRA Analysis

The video and audiotapes of the groups' discourse throughout their exploration of the computer simulations were transcribed verbatim. The discourse was segmented and coded by conversational turns (i.e., changes in the speaker). Three sets of codes were developed and applied to investigate students' collaborative learning through different lenses: collaborative discourse, epistemic practices, and SBF content (Liu 2008).

The *collaborative discourse* coding examined cognitive and metacognitive processes underlying the group's discourse and facilitators' roles. There are three major discourse subcategories: cognitive processing, metacognitive processing, and teacher facilitation. Details are presented in Table 4.1 and further elaborated in Liu (2008) and Hmelo-Silver et al. (2009).

The second coding scheme was developed to capture the characteristics of *epistemic practices* (i.e., the practices embodying ways of scientific thinking and how learners work on knowledge construction task, see in Duschl and Osborne (2002) to build understanding (Table 4.2). The coding categories present a set of discursive practices for generating and evaluating knowledge (Liu 2008; Hmelo-Silver et al. 2009).

The third scheme coded for *SBF content* and was used to investigate the extent to which students talked about structures, behaviors, and functions (SBF; Hmelo-Silver et al. 2009). This allowed us to examine how the students talked about content as they engaged in their exploration of the simulation, particularly in the context of how we had structured the instruction. The instructional intervention was organized to help the students learn to use SBF as a way of thinking about systems. We coded for structures in statements that are focused on the 'what's' of the system such as "What is the red dot?" as students were trying to determine what a simulation object was. We coded for behaviors when students referred to how the system worked. We coded statements as functions if they referred to the roles outputs of different parts of the system. An example function statement is "the fish excrete ammonia."

To understand how the collaboration unfolded, we turn to the CORDTRA diagrams in Figs. 4.6 and 4.7. These diagrams focus on one of harder aspects of the task—student discourse as they tried to make sense of the nitrification process while working with the microlevel NetLogo simulation. In group 1 (the high achieving group shown in Fig. 4.6), the students talked in terms of observations

Table 4.1 Definitions for collaborative coding categories

Categories	Definitions	Examples
Cognitive process		
Fact question	Questions asked with a purpose to obtain factual information	"What is the yellow stuff?"
Explanation question	Questions asked with a purpose to obtain cause-effect information	"Why is water qualify dropping?"
Confirm question	Questions asked to make sure one gets the shared information	"The males couldn't wait to make more fish so they what?"
Directing statement	Demanding statement for an ongoing activities	"Change the water now."
Agree	Explicit express of acceptance of other's ideas	"Okay I guess that makes clear sense."
Disagree	Expressing express of rejection of other's ideas	"No. This is not true."
Share knowledge	Share information with other members in the group	"I have fish, plants, bacteria1, bacteria2, ammonia, nitrite and nitrate."
Describe observation	Descriptions of what is observed in the simulations	"Now there are no more male fish"
Retrieve prior knowledge	Making connections to one's previously perceived knowledge or experiences	"We know that there is bacteria inside the water that eats the bad bacteria."
Generate theory	Statement of a hypothetical proposal	"When there were more female fish they ate all the smaller fish and then died."
Paraphrase	Rewording other's statements	"Okay so when there were more female fish they ate the smaller fish and died of old age."
Warranted claim	Statements to provide ground for an idea	"Well we are looking at the chart and it tells how ammonia, the bacteria turns it into nitrate. Doesn't it kind of prove that the stuff in the back is bacteria then"
Identify cognitive conflict	Realizing the discrepancies in one's or the group's reasoning	"Because the model we have is that when there are more female fish they eat the smaller fish and then they died of old age. But then they are eating the smaller fish and none of them are dying of old age."
Off-topic talking	Statement unrelated to the learning target	"Can I borrow your pen?"
Metacognitive process		
Plan	Defining the learning goals	"Okay we have to figure out what they do."

(continued)

Table 4.1 (continued)

Categories	Definitions	Examples
Monitor	Reflecting on the learning process to keep track of the conceptual understanding	"We haven't explain how they keep a balance?"
Review	Looking back on the strategies (e.g., designing experiments, running simulations) that lead to knowledge construction	"Well we tried to take away the plants … and then nothing even happened"
Evaluate	Judging the effectiveness of learning strategies	"Using one fish for each gender helped to find out which gender lives longer."
Facilitators' roles		
Educational statement	Statements related to the learning content and strategies	"You need to move on to the next question."
Performance statement	Statements related to class management and students' performance	"Try to look at the hypermedia. Maybe you will get some information there."
Open question	Questions seeking an elaborated answer or explanation	"How do you know the water quality has decreased?"
Closed questions	Questions seeking a short and factual answer	"Are all of those bad for water quality?"

From Liu (2008)

about colored dots (which represent different chemicals: ammonia, nitrite, and nitrate) and patches (which represent two types of bacteria) during the first half of the discussion; in second half, the students began to talk more about what those patches represent. Further, Fig. 4.6 demonstrates that the group conversation shifted across all SBF levels and initially shifting largely between structures and behaviors, as they were trying to understand what was happening in the simulation. These connections between structures, behaviors, and functions are not normally made by novices. Inspection of Fig. 4.6 also shows that associated with this shifting, the students often engaged in exchanging knowledge, warranting claims and in the middle part of the discourse, with designing experiments.

The CORDTRA diagram for the low achieving group is shown in Fig. 4.7. In contrast to the high achieving group, this group began their discussion at the structural and behavioral level. They engaged in some discussion of function in the middle, but then ended with continued discussion of experimental designs, which were not driven by explicit goals and were *not* associated with shifting between SBF levels and this group did not often discuss function. This resulted in a discussion of behaviors that created a description rather than an explanation. Although there was an increase in explanation questions over time, these students were still asking many fact-oriented questions about what they directly observed. Similar to

Table 4.2 Definitions for epistemic practice coding categories

Categories	Definitions	Examples
Basic knowledge construction	Superficial meaning making practice without reasoning or supporting evidence	"What is the yellow? Yeah, I think is food or is that like dirt?"
Observe	Practices of observing phenomena on the computer screen	"Wow! Look it, it went down real quick."
Predict	Practices aiming to propose predicting result of a simulation	"And if you increase it to 2,000 they'll die more quicker."
Design experiment	Designing a simulating experiment to test hypotheses	"How about we if put this, and this all the way down to zero? And put this thing on the top?"
Check knowledge validity	Examine the consistency or accountability of constructed knowledge by taking several experimental trials.	"No, see this number is like the same, whatever this corresponds to this. It's still 8. Ammonia and saturated in nitrite. But 82 and 75…it adds up to the same number."
Coordinate theory-evidence	Practices entailing using theories to explain data and using data to evaluate theories	"So the plants absorb nitrite because the yellow disappeared. Nitrite, which….comes from nitrate. Nitrate with an A, nitrate comes from nitrite, nitrite comes from ammonia, from the bacteria, the white went in and went through the patch."
Modify knowledge	Making a change in previously constructed knowledge	"No, the patch is not fish. It is bacteria."
Exchange knowledge	Explicit articulation of one's knowledge to others.	"So you are saying fish excrete ammonia to become nitrite."
Give feedback	Providing evaluative responses to other's statements or actions	"Yes, you are right. The red dots disappeared."
Scaffold	Applying purposeful strategies to support other's understanding (subjected to teacher's conversational turns only)	"So what does that explain about different kinds of models?"

From Liu (2008)

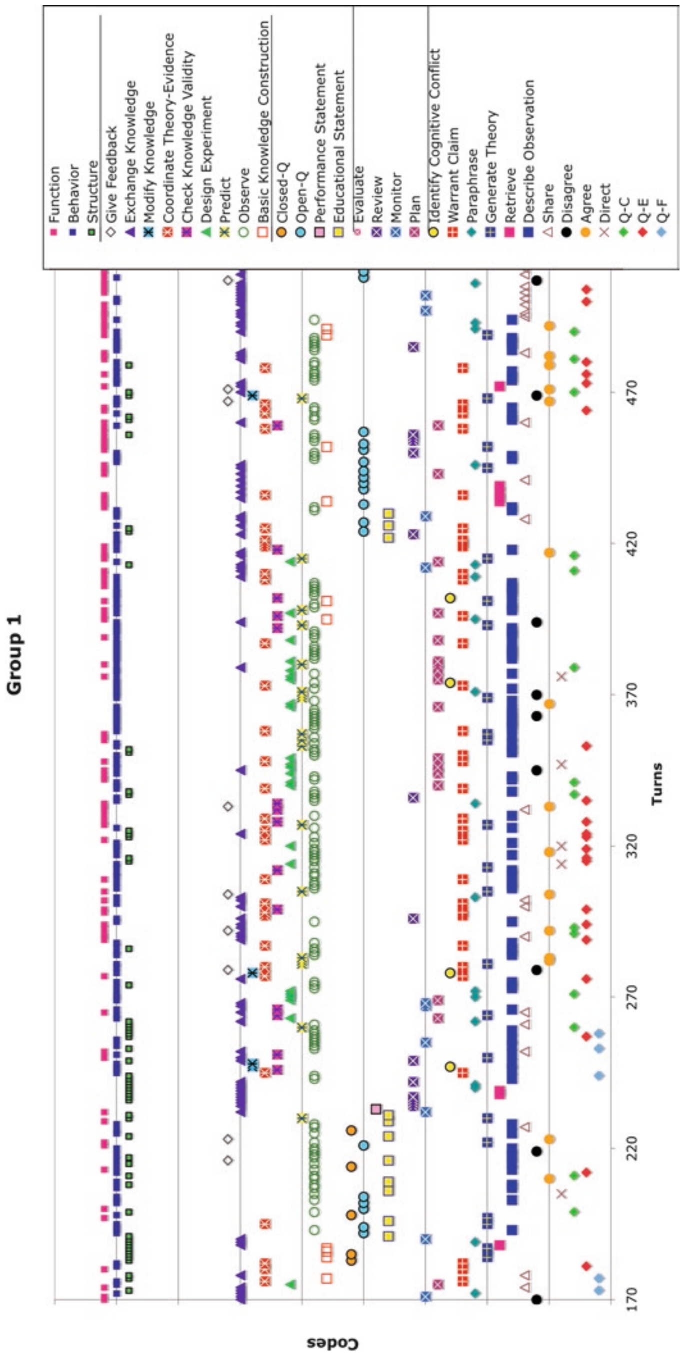

Fig. 4.6 High Achieving group working with simulation (Hmelo-Silver et al. 2009)

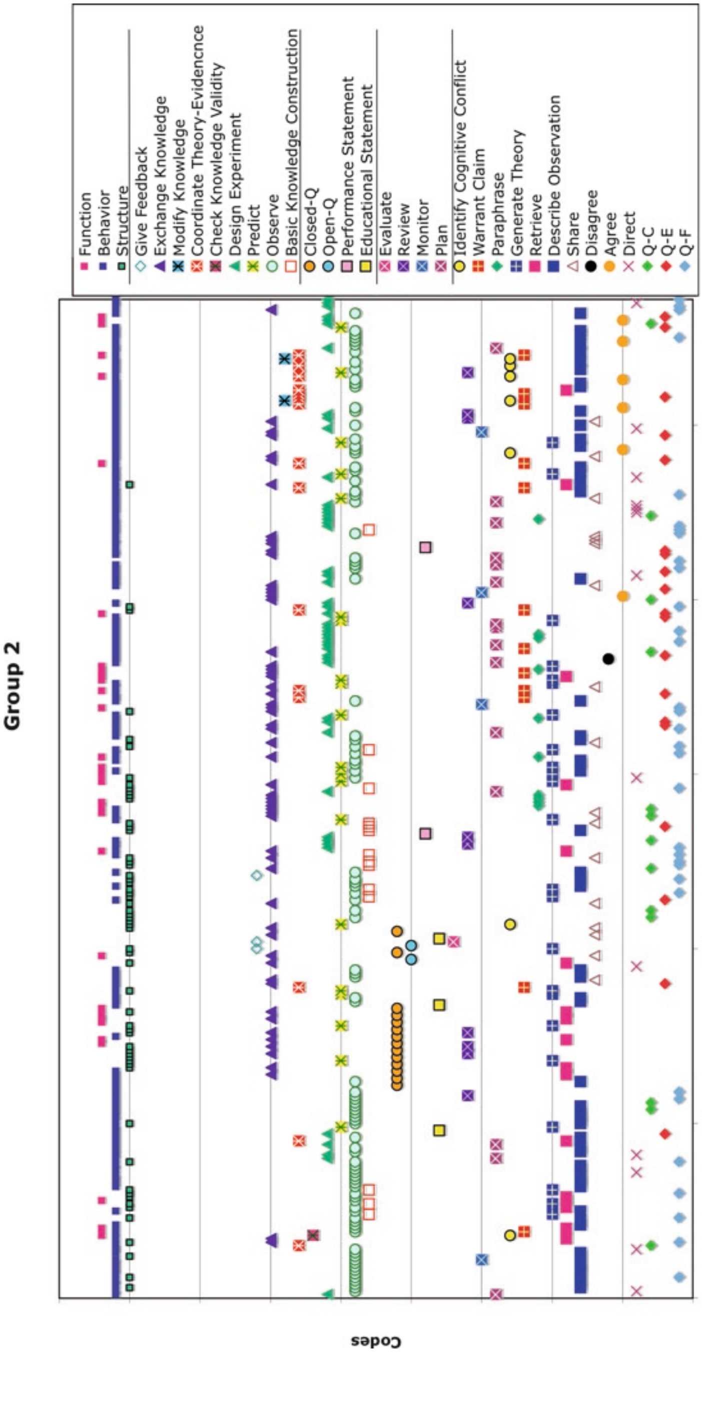

Fig. 4.7 Low Achieving group working with simulation (Hmelo-Silver et al. 2009)

the high achieving group, the times when this group shifted between SBF levels was associated with knowledge exchange. Although the discourse included many instances of warranting claims and identifying cognitive conflicts, a large portion of the discourse stayed at the behavioral level. The group was focusing on manipulating the simulations and observing relationships but not getting to the functional aspects that would let them construct an explanation. Without comparing the different aspects of the discourse including the SBF content, the discourse and the epistemic practice features, it is unlikely that we would see the whole picture regarding the quality of the group collaboration. The CORDTRA diagram helps us to see the relations between SBF topics and the discourse and epistemic features, which leads to a thorough understanding of the collaborative process.

4.5 Discussion

Like others in the CSCL community, we are concerned with how students negotiate meaning in collaborative groups and how this learning is mediated by technology (Suthers 2006). Visual representations can be useful in trying to see patterns in complex CSCL data and visual representations integrating time can enable us to see patterns in data that might be obscured if these data were aggregated into frequencies. Representational tools are needed to allow researchers to understand how learning happens among social interactions and use of computer tools.

In particular, we have found CORDTRA diagrams to be a useful tool to visually represent complex and dynamic learning data. In Example 1, the CORDTRA results helped us generate hypotheses related to understanding group collaboration and the use of online tools. For example, we found that in effective groups, learners (a) build on each other's ideas, (b) understand the task that they are engaging in at the outset, (c) process the information they gather in their self-directed learning, and (d) use the right resources at the right time. Future analyses in a range of contexts will be needed to see if this pattern holds. In Example 2, the CORDTRA analysis was particularly helpful in understanding the importance of students cycling between structure, behavior, and function levels as they construct descriptions or explanations. Explanations seemed to occur as students talked about function while coordinating theory and evidence, particularly late in the work in the simulation.

It is clear that in CSCL environments, learners may need additional support from teachers or scaffolds built into the technology to promote opportunities for effective collaboration. The detailed CORDTRA analyses can enable us to consider the relation between collaborative learning and tool use and provide suggestions for how this support might be accomplished. Further, CORDTRA can quickly enable us to highlight where the interesting activity is in CSCL environments that is worthy of further analysis. This can be helpful when analyzing the learning environments for effectiveness and user-friendliness. Can also help both the facilitators and researchers to focus on the appropriate parts of the activity that need to be improved either while teaching or during the post-teaching analyses. Accomplishing this, however, would require automating construction of CORDTRA diagrams so that they can be available in a timely fashion. More effort needs to be expended on automatic construction of CORDTRA diagrams to enable further understanding of dynamic processes occurring in CSCL environments. In conclusion, representations such as CORDTRA can enable both researchers and instructors to better understand complex learning environments, because a "diagram can (sometimes) be worth 10,000 words" (Larkin and Simon 1987, p. 65).

References

Arrow, H., Henry, K. B., Poole, M. S. S. W., & Moreland, R. (2005). Traces, trajectories, and timing: the temporal perspective on groups. In M. S. Poole & A. B. Hollingshead (Eds.), *Theories of small groups: interdisciplinary perspectives* (pp. 313–367). CA: Sage.

Avouris, N., Dimitracopoulou, A., & Komis, V. (2003). On analysis of collaborative problem solving: an object-oriented approach. *Computers in Human Behavior, 19*, 147–167.

Chernobilsky, E., Hmelo-Silver, C. E., & DelMarcelle, M. (2003). *Collaborative discourse, tools, and activity in online problem-based learning*. Chicago: Paper presented at the Annual meeting of the American Educational Reseach Association.

Chinn, C. A., & Anderson, R. C. (2000). The structure of discussions that promote reasoning. *Teachers College Record, 100*, 315–368.

Chiu, M. M. (2008). Flowing towards correct contributions during problem solving: a statistical discourse analysis. *Journal of the Learning Sciences, 17*, 415–463.

Cole, M., & Engeström, Y. (1993). A cultural-historical approach to distributed cognition. In G. Salomon (Ed.), *Distributed cognitions: psychological and educational considerations*. New York: Cambridge University Press.

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.

De Laat, M. F., & Lally, V. (2003). Complexity, theory, and praxis: researching collaborative learning and tutoring processes in a networked learning community. *Instructional Science, 31*, 7–39.

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

Derry, S. (2006). STEP as a case of theory-based web course design. In A. M. O'Donnell, C. E. Hmelo-Silver, & G. Erkens (Eds.), *Collaborative reasoning, learning and technology*. Mahwah: Erlbaum.

Derry, S. J., Hmelo-Silver, C. E., Nagarajan, A., Chernobilsky, E., & Beitzel, B. (2006). Cognitive transfer revisited: can we exploit new media to solve old problems on a large scale? *Journal of Educational Computing Research, 35*, 145–162.

Duschl, R., & Osborne, J. (2002). Argumentation and Discourse Processes in Science Education. *Studies in Science Education, 38*, 39–72.

Engeström, Y. (1999). Activity theory and individual and social transformation. In Y. Engström, R. Miettinen, & R. Punamaki (Eds.), *Perspectives on activity theory* (pp. 19–38). New York: Cambridge University Press.

Engle, R. A., & Conant, F. R. (2002). Guiding principles for fostering productive disciplinary engagement: explaining an emergent argument in a community of learners classroom. *Cognition and Instruction, 20*, 399–484.

Fischer, F., Bruhn, J., Gräsel, C., & Mandl, H. (2002). Fostering collaborative knowledge construction with visualization tools. *Learning and Instruction, 12*, 213–232.
Goel, A. K., de Silva Garza, Gomez, Grué, N., Murdock, J. W., Recker, M. M., & Govinderaj, T. (1996). Towards designing learning environments -I: exploring how devices work. In C. Fraisson, G. Gauthier, & A. Lesgold (Eds.), *Intelligent tutoring systems: lecture notes in computer science* (pp. 42–52). NY: Springer.
Hmelo, C. E., Guzdial, M., & Turns, J. (1998). Computer support for collaborative learning: learning to support student engagement. *Journal of Interactive Learning Resarch, 9*, 107–130.
Hmelo-Silver, C. E. (2003). Analyzing collaborative knowledge construction: multiple methods for integrated understanding. *Computers & Education, 41*, 397–420.
Hmelo-Silver, C. E., Chernobilsky, E., & Jordan, R. (2008). Understanding collaborative learning processes in new learning environments. *Instructional Science, 36*, 409–430.
Hmelo-Silver, C. E., Katic, E., Nagarajan, A., & Chernobilsky, E. (2007a). Soft leaders, hard artifacts, and the groups we rarely see: using video to understand peer-learning processes. In R. Goldman, R. D. Pea, B. J. S. Barron, & S. J. Derry (Eds.), *Video research in the learning sciences* (pp. 255–270). Mahwah: Erlbaum.
Hmelo-Silver, C. E., Liu, L., Gray, S., Finkelstein, H., & Schwartz, R. (2007b). Enacting things differently: Using NetLogo models to learn about complex systems. Paper presented at the Biennial meeting of European Association for Research on Learning and Instruction.
Hmelo-Silver, C. E., Liu, L., & Jordan, R. (2009). Visual representation of a multidimensional coding scheme for understanding technology-mediated learning about complex natural systems. *Research and Practice in Technology-enhanced Learning Environments, 4*, 253–280.
Hmelo-Silver, C. E., Marathe, S., & Liu, L. (2007c). Fish swim, rocks sit, and lungs breathe: expert-novice understanding of complex systems. *Journal of the Learning Sciences, 16*, 307–331.
Janssen, J., Erkens, G., Kanselaar, G., & Jaspers, J. (2007). Visualization of participation: does it contribute to successful computer-supported collaborative learning. *Computers & Education, 49*, 1037–1065.
Jeong, A., Clark, D. B., Sampson, V. D., Menekse, M. this volume. Sequential analysis of scientific argumentation in asynchronous online discussion environments. In S. Puntambekar, G. Erkens, & C. E. Hmelo-Silver (Eds.) *Analyzing interactions in CSCL: Methods, approaches, and issues* (pp. 207–233). Springer.
Kumpulainen, K., & Mutanen, M. (1999). The situated dynamics of peer group interaction: an introduction to an analytic framework. *Learning and Instruction, 9*, 449–473.
Lai, M., & Law, N. (2006). Peer scaffolding of knowledge building through collaborative groups with differential learning experiences. *Journal of the Educational Computing Research, 35*, 123–144.
Lajoie, S. P., Garcia, B., Berdugo, G., Márquez, L., Espíndola, S., & Nakamura, C. (2006). The creation of virtual and face-to-face learning communities: an international collaboration experience. *Journal of the Educational Computing Research, 35*, 163–180.
Larkin, J. H., & Simon, H. A. (1987). Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science, 11*, 65–99.
Liu, L. (2008). *Trajectories of collaborative scientific conceptual change: middle school students learning about ecosystems in a CSCL environment*. New Brunswick: Unpublished Dissertation.
Luckin, R. (2003). Between the lines: documenting the multiple dimension of computer-supported collaborations. *Computers & Education, 41*, 379–396.
Luckin, R., Plowman, L., Laurillard, D., Stratfold, M., Taylor, J., & Corben, S. (2001). Narrative evolution: learning from students' talk about species variation. *International Journal of Artificial Intelligence in Education, 12*, 100–123.
Martinez, A., Dimitriadis, Y., Rubia, B., Gomez, E., & de la Fuente, P. (2003). Combining qualitative evaluation and social network analysis for the study of classroom social interactions. *Computers & Education, 41*, 353–368.

McGrath, J. E., Arrow, H., & Berdahl, J. L. (2000). The study of small groups, past, present, and future. *Personality and Social Psychology Review, 4*, 95–105.
Mercer, N. (2008). The seeds of time: why classroom dialogue needs a temporal analysis. *Journal of the Learning Sciences, 17*, 33–59.
Nurmela, K., Lehtinen, E., & Palonen, T. (1999). Evaluating CSCL log files by social network analysis. In C. M. Hoadley & J. Roschelle (Eds.), *Proceedings of the 1999 conference on computer support for collaborative learning* (pp. 54–66). Mahwah: Erlbaum.
Palincsar, A. S. (1998). Social constructivist perspectives on teaching and learning. *Annual Review of Psychology, 45*, 345–375.
Reimann, P. (2007). Time is precious: why process analysis is essential for CSCL (and can also help to bridge between experimental and descriptive methods). In C. A. Chinn, G. Erkens, & S. Puntambekar (Eds.), *Proceedings of CSCL 2007: mice, minds, and society* (Vol. 8, pp. 605–614). New Brunswick: International Society of the Learning Sciences.
Roschelle, J. (1996). Learning by collaborating: convergent conceptual change. In T. D. Koschmann (Ed.), *CSCL: theory and practice of an emerging paradigm* (pp. 209–248). Mahwah: Erlbaum.
Rummel, N., Meier, A., Spada, H., & Kahrimanis (2011). Analyzing collaborative interactions across domains and settings: An adaptable rating scheme. In S. Puntambekar, G. Erkens, & C. E. Hmelo-Silver (Eds.), *Analyzing interactions in CSCL: Methodology, approaches, and issues*. Springer.
Rummel, N., & Spada, H. (2005). Learning to collaborate: an instructional approach to promoting collaborative problem solving in computer-mediated settings. *Journal of the Learning Sciences, 14*, 201–241.
Schümmer, T., Strijbos, J. W., & Berkel, T. (2005). Measuring group interaction during CSCL. In T. Koschmann, D. Suthers, & T. W. Chan (Eds.), *Computer supported collaborative learning 2005: the next 10 years!* (pp. 567–576). Mahwah: Lawrence Erlbaum Associates.
Schwartz, D. L., & Bransford, J. D. (1998). A time for telling. *Cognition and Instruction, 16*, 475–522.
Stahl, G. (2006). Supporting group cognition in an online math community: a cognitive tool for small-group referencing in text chat. *Journal of the Educational Computing Research, 35*, 103–122.
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.
Strom, D., Kemeny, V., Lehrer, R., & Forman, E. (2001). Visualizing the emergent structure of children's mathematical argument. *Cognitive Science, 25*, 733–774.
Suthers, D. D. (2006). Technology affordances for intersubjective meaning making. *International Journal of Computer Supported Collaborative Learning, 1*, 315–337.
Suthers, D. D., Dwyer, N., Medina, R., & Vatrapu, R. (2009). Exposing interactional processes in online learning. In K. Kumpulainen, C. E. Hmelo-Silver, & M. César (Eds.), *Investigating classroom interaction: methodologies in action*. Rotterdam: Sense.
Suthers, D. D., & Hundhausen, C. D. (2003). An experimental study of the effects of representational guidance on collaborative learning processes. *Journal of the Learning Sciences, 12*, 183–218.
Suthers, D., & Medina, R. (this volume). Tracing interaction in distributed collaborative learning. In S. Puntambekar, G. Erkens, & C. E. Hmelo-Silver (Eds.), *Analyzing interactions in CSCL: Methods, approaches, and issues* (pp. 341–366). Springer.
Wilensky, U., & Reisman, K. (2006). Thinking like a wolf, sheep, or firefly: learning biology through constructing and testing computational theories– an embodied modeling approach. *Cognition and Instruction, 24*(2), 171–209.
Zemel, A., Xhafa, F., & Cakir, M. (2007). What's in the mix? Combining coding and conversation analysis to investigate chat-based problem-solving. *Learning and Instruction, 17*, 405–415.

Chapter 5
How to Study Group Cognition

Gerry Stahl

Abstract Understanding how a collaborative group as a whole constructs knowledge through joint activity in a CSCL setting is what sets the research field of CSCL apart from other approaches to the study of learning. Successful collaboration involves not only the incorporation of contributions of individuals into the group discourse, but also the effort to make sure that participating individuals understand what is taking place at the group level. The contributions of individuals to the group and of understandings from the group to the individuals cannot be studied by analyses at the individual unit of analysis, but only by studying the interactions at the group level. The group knowledge construction process synthesizes innumerable resources from language, culture, the group's own history, individual backgrounds, relevant contexts and the sequential unfolding of the group discourse in which the individuals participate. Although the group process is dependent upon contributions and understanding of individuals, their individual cognition is essentially situated in the group process. Group cognition is the science of cognitive processes at the group unit of analysis. These group processes—such as the sequential flow of proposals, questioning, building common ground, maintaining a joint problem space, establishing intersubjective meanings, positioning actors in evolving roles, building knowledge collaboratively and solving problems together—are not analyzable as individual behaviors. This chapter will describe how the Virtual Math Teams project was designed as a prototypical CSCL environment in which the relevant resources and interactions could be recorded for the micro-analytic study of group cognition.

5.1 The Need for a New Science of Group Cognition

The design of software to support group work, knowledge building and problem solving should be built on the foundation of an understanding of the nature of group interaction. This chapter argues that previous research in CSCW and CSCL is based

G. Stahl (✉)
College of Information Science and Technology, Drexel University,
3141 Chestnut St, Philadelphia, USA
e-mail: Gerry@GerryStahl.net

S. Puntambekar et al. (eds.), *Analyzing Interactions in CSCL,* Computer-Supported Collaborative Learning Series, DOI 10.1007/978-1-4419-7710-6_5,

on an ad hoc collection of incommensurable theories, which are not grounded in an explicit investigation of group interaction. What is needed is a science of group interaction focused on the group level of description to complement psychological theories of individuals and social theories of communities.

Preparing for a new science requires three major undertakings:

(a) The domain of the science must not only be defined, it must be explored and captured in the form of a data corpus.
(b) Methods for analyzing the data must be selected, adapted, refined and mastered.
(c) Analytic findings must be organized in terms of a framework of theoretical conceptualizations.

After discussing the need for a new science of group interaction, this chapter describes a research project that successfully approached these tasks by:

(a) Creating a synchronous online service in which small groups of students engaged in problem-solving work in mathematics,
(b) Conducting chat interaction analysis of a number of case studies from the data recorded in that service and
(c) Conceptualizing some of the features of the small-group interactions that were observed.

The focus on small groups was originally motivated by the realization that CSCW and CSCL are fundamentally different from other domains of study (Stahl 2002). They take as their subject matter *cooperative* working or *collaborative* learning, that is, what takes place when small groups of workers or students engage together in cognitive activities like problem solving or knowledge building (Koschmann 1996; Stahl 2006b, ch. 11). On a theoretical level, CSCW and CSCL are strongly oriented toward Vygotsky (1930/1978), who stressed that learning and other higher psychological processes originally take place socially, intersubjectively. (Piaget 1985), too, pointed to inter-subject processes like conflicting perspectives as a fundamental driver for creativity and cognitive development. Despite this powerful insight, Vygotsky, Piaget and their followers generally maintain a psychological focus on the individual mind in their empirical studies and do not systematically investigate the intersubjective phenomena of small-group interaction.

A science of group interaction would aim to unpack what happens at the small-group unit of analysis (Stahl 2004a). Thus, it would be particularly relevant for CSCW and CSCL, but may not be as directly applicable to other forms of working or learning, where the individual or the community level predominates. As a science of the group, it would complement existing theories of working, learning and cognition, to the extent that they focus either on the individual or the community or that they reduce group phenomena to these other levels of description.

In the chapters of *Studying Virtual Math Teams* (Stahl 2009b) and of *Group Cognition* (Stahl 2006b), my colleagues and I have reviewed some of the research literature on small-group learning, on small-group processes and on collaborative mathematics. We have noticed that small-group studies generally look for quantitative correlations among variables—such as the effect of group size on measures of participation—rather than trying to observe group knowledge-building processes. Studies of small-group processes from psychology, sociology and other social sciences also tend to focus on non-cognitive aspects of group process or else attribute all cognition to the individual minds rather than to group processes. There are some notable exceptions; in particular, we viewed (Barron 2000, 2003; Cohen et al. 2002; Sawyer 2003; Schwartz 1995) as important preliminary studies of group cognition. However, even theories that seem quite relevant to our concerns, like distributed cognition (Hutchins 1996), actor-network theory (Latour 2007), situated cognition (Lave and Wenger 1991), ethnomethodology (Garfinkel 1967) and activity theory (Engeström 1987) adopt a different focus, generally on interaction of individuals with artifacts rather than among people. In particular, recent commentaries on situated cognition (Robbins and Aydede 2009) and distributed cognition (Adams and Aizawa 2008) frame the issues at the individual level, even reducing all cognitive phenomena to neural phenomena. At the other extreme, social theories focus on community phenomena like division of labor, apprenticeship training, linguistic structure, laboratory organization. For all its insight into small-group interaction and its analysis, ethnomethodology maintains a sociological perspective. Similarly, even when activity theory addresses the study of teams—in the most detail in Chap. 6 of (Engeström 2008)—it is mostly concerned with the group's situation in the larger industrial and historic context; rather than analyzing how groups interactionally build knowledge it paraphrases how they deal politically with organizational management issues. These theories provide valuable insights into group interaction, but none of them thematizes the small-group level as a domain of scientific study. As sciences, these are sciences of the individual or of the society, not of the collaborative group.

Each of the three levels of description is populated with a different set of phenomena and processes. For instance, *individuals* interpret recent postings and design new postings in response, the *group* constructs, maintains and repairs a joint problem space and the *community* evolves its shared methods of social organization. The description of the individual level is the province of psychology; that of the community is the realm of sociology or anthropology; *the small-group level has no corresponding science*.

A science of group interaction would take its irreducible position between the psychological sciences of the individual and the social sciences of the community—much as biology analyzes phenomena that are influenced by both chemicals and organisms without being reducible to either. The science of group interaction would fill a lacuna in the multi-disciplinary work of the human sciences. This science would not be primarily oriented toward the "low level" processes of groups, such as mechanical or rote behaviors, but would be concerned with the accomplishment of creative intellectual tasks. Intellectual teamwork, knowledge work and knowledge-building activities would be prototypical objects of study. The focus would be on group cognition.

The bifurcation of the human sciences into individual and societal creates an irreconcilable opposition between individual creative freedom and restrictive social institutions. A science of group cognition would flesh out the concept of structuration, demonstrating with detailed analyses of empirical data how group interactions can mediate between individual behavior and social practices (Stahl 2009b, ch. 11).

The term *group cognition* does not signify an object or phenomenon to analyze like brain functions or social institutions (Stahl 2004b). It is a proposal for a new science or focus within the human sciences. It hypothesizes:

> When small groups engage in cooperative problem solving or collaborative knowledge building, there are distinctive processes of interest at the individual, small-group and community levels of analysis, which interact strongly with each other. The science of group cognition is the study of the processes at the small-group level.

The science of group cognition is a human science, not a predictive science like chemistry, nor a predominantly quantitative one like physics. It deals with human meanings in unique situations, necessarily relying upon interpretive case studies and descriptions of inter-personal processes.

Processes at the small-group level are not necessarily reducible to processes of individual minds, nor do they imply the existence of some sort of group mind. Rather, they may take place through the weaving of semantic and indexical references within a group discourse. The indexical field (Hanks 1992) or joint problem space (Teasley and Roschelle 1993) co-constructed through the sequential interaction of a group (Cakır et al. 2009) has the requisite complexity to constitute an irreducible cognitive structure in its own right. Cognitive science broadened the definition of "cognition" beyond an activity of human minds in order to include artificial intelligence of computers. What counts as cognitive is now a matter of computational complexity. Anything that can compute well enough to play chess or prove theorems can be a cognitive agent—whether they are a person, computer or collaborative small group (Stahl 2005).

A science of group cognition is timely and relevant, as indicated by the rise of the CSCW and CSCL fields. The 21st century will increasingly rely on small groups—due to networked computers creating new means of group intellectual production, with the power to overcome the limitations of the individual mind. The dominance of the individual in production and in science was part of the larger epochal trend of industrialization. Now forces of instantaneous communication, globalization and ecological crisis seem to be bringing about a transformation of that historic trend, resulting in the rising prominence of the small group as an important mediator between the isolated individual and an increasingly abstract society. The small group is becoming an effective new form in the social relations of intellectual production.

Having motivated the development of a science of group cognition as future work, let us see how the Virtual Math Teams (VMT) Project may have begun to prepare the way. We start with how the VMT world of online collaboration was constructed as an object of study. In this chapter, we describe the nature of science that is projected by the analyses of the VMT Project, leaving the specific case studies of group cognition phenomena to our other publications. For instance, we have elsewhere described the sequential flow of proposals (Stahl 2006d), questioning (Zhou 2009), building common ground (Stahl 2006c), maintaining a joint problem space (Sarmiento and Stahl 2008), establishing intersubjective meaning (Stahl 2007), positioning actors in evolving roles (Charles and Shumar 2009), building knowledge collaboratively (Stahl 2006a) and solving problems together (Cakır et al. 2009), using excerpts from VMT sessions. While other CSCL researchers have found these analyses interesting, they often raise questions about our scientific method. Therefore, in this chapter we have focused on describing the distinctive methodological approach of a science of group cognition.

5.2 Designing a Testbed for Studying Group Cognition

The first step in our design-based research process was to start simply and see what issues came up. We had seen in face-to-face case studies that there were problems with (a) recording and transcribing the verbal interaction, (b) capturing the visual interaction and (c) knowing about all the influences on the interaction. We decided to form groups of students who did not know each other and who only interacted through text chat. Students were recruited through the Math Forum at Drexel University, an established online resource center. We used AIM, AOL's Instant Messaging system, which was freely available and was already familiar to many students. We included a researcher in the chat room with each small group of students. The facilitator told the students their math task, dealt with any technical difficulties, posted drawings from the students on a web page where they could be seen by all the students, notified the group when the session was over and saved an automatically generated log of the chat. In this way, we obtained a complete and objective log of the interaction, captured everything that the students shared on their computers and excluded any unknown influences from affecting the group interaction.

The issue of including everything affecting the interaction is a subtle issue. Of course, the interaction is influenced by the life histories, personalities, previous knowledge and physical environment of each student. A student may have windows other than AIM open on the computer, including Internet browsers with math resources. A student may be working out math problems on a piece of paper next to the computer. Also, a student may leave the computer for some time to eat, listen to music, talk on the phone, and so on without telling anyone in the chat. In such ways, we do not have information about everything involved in a particular student's online experience. We do not even know the student's gender or age. We do not know if the student is shy or attractive, speaks with an accent or stutters. We do not know if the student usually gets good grades or likes math. We do not know what the student is thinking or feeling. We only know that the students are in an approximate age group and academic level—because we recruited them through teachers. However, the VMT Project is only concerned with analyzing the interaction at the *group unit of analysis*. Notice that the things that are unknown to us as researchers are also unknown to the student group as a whole. The students do not know specifics about each other's background or activities—except to the extent that these specifics are brought into the chat. If they are mentioned or referenced in the chat, then we can be aware of them to the same extent as are the other students.

The desire to generate a complete record for analysis of everything that was involved in a team's interaction often conflicted with the exploration of technology and service design options. For instance, we avoided speech-based interaction (VOIP, Skype, WIMBA) and support for individual work (e.g., whiteboards for individual students to sketch ideas privately), because these would complicate our review of the interactions. We tried to form teams that did not include people who knew each other or who could interact outside of the VMT environment.

In addition to personal influences, the chat is responsive to linguistic and cultural matters. Of course, both students and researchers must know English to understand the chats. In particular, forms of English that have evolved with text chat and cell-phone texting have introduced abbreviations, symbols and emoticons into the online language. The linguistic subculture of teenagers also shows up in the VMT chats. An interdisciplinary team of researchers comes in handy for interpreting the chats. In our case, the research team brought in experience with online youth lingo based on their backgrounds as Math Forum staff, teachers or parents.

The early AIM chats used simple math problems, taken from standardized math tests and Math Forum Problems-of-the-Week. One experiment to compare individual and group work used problems from a standardized multiple-choice college-admissions test (Stahl 2009a). These problems had unique correct answers. While these provided a good starting point for our research, they were not well suited for collaborative knowledge building. Discourse around them was often confined to seeing who thought they knew the answer and then checking for correctness. For the VMT Spring Fests in 2005, 2006 and 2007, we moved to more involved math topics that could inspire several hours of joint inquiry.

Even with straight-forward geometry problems, it became clear that students needed the ability to create, share and modify drawings within the VMT environment. We determined that we needed an object-oriented draw program, where geometric objects could be manipulated (unlike a pixel-based paint program). We contracted with the developers of ConcertChat to use and extend their text chat and shared whiteboard system, which is now available in Open Source. This system included a graphical referencing tool as well as social awareness and history features (Mühlpfordt and Stahl 2007). In order to help students find desirable chat rooms and to preserve team findings for all to see, we developed the VMT Lobby and integrated a Wiki with the Lobby and chat rooms (Stahl 2008b). Gradually, the technology and the math topics became much more complicated in response to the needs that were revealed when we analyzed the trials of the earlier versions of the VMT service. As the system matured, other research groups began to use it for their own trials, with their own math topics, procedures, analytic methods or even new technical features. These groups included researchers from Singapore (Wee and Looi 2009), Rutgers (Powell and Lai 2009), Hawai'i (Medina et al. 2009), Romania (Trausan-Matu and Rebedea 2009) and Carnegie-Mellon (Cui et al. 2009).

The evidence for the adequacy of a testbed for design-based research lies in the success of the analyses to reveal how the prototyped environment is working at each iteration and to provide ideas based on problems encountered by users to drive the design further. Therefore, we now turn to the analyses of interaction in the virtual math teams to see if the testbed produced adequate data for understanding group cognition in this context.

5.3 Studying Group Cognition

The approach to chat interaction analysis that emerged in the VMT Project will be discussed in this section in terms of a number of issues (which correspond to general issues of most research methodologies, as indicated in parentheses):

5.3.1 Group Cognition in a Virtual Math Team (Research Question)

Learning—whether in a classroom, a workplace or a research lab—is not a simplistic memorization or storage of facts or propositions, as traditional folk theories had it. The term *learning* is a gloss for a broad range of phenomena, including: the development of tacit skills, the ability to see things differently, access to resources for problem solving, the discursive facility to articulate in a new vocabulary, the power to explain, being able to produce arguments or the making of new connections among prior understandings (Stahl and Herrmann 1999). We can distinguish these phenomena as taking place within individual minds, small-group interactions or communities of practice. The analysis of learning phenomena at these various levels of analysis requires different research methodologies, appropriate to corresponding research questions. The VMT Project was intended to explore the phenomena of group cognition and accordingly pursued the research question:

> How does learning take place in small groups, specifically in small groups of students discussing math in a text-based online environment? *What are the distinctive mechanisms or processes that take place at the small-group level of description* when the group is engaged in problem-solving or knowledge-building tasks?

While learning phenomena at the other levels of analysis are important and interact strongly with the group level, we have tried to isolate and make visible the small-group phenomena and to generate a corpus of data for which the analysis of the group-level interactions can be distinguished from the effects of the individual and community levels.

The methods used to gather and analyze one's data should be appropriate to one's research question. To support such research, one must generate and collect data that are adequate for the selected kinds of analysis. Because we were interested in the group processes that take place in virtual math teams, we had to form teams that could meet together online. In the Spring Fests, students had to be able to come back together in the same teams on several subsequent occasions. The VMT environment had to be instrumented to record all messages and activities that were visible to the whole team in a way that could be played back by the analysts. The math problems and the feedback to the teams had to be designed to encourage the kinds of math discussions that would demonstrate processes of group cognition, such as formulating questions and proposals, coordinating drawings and textual narratives, checking proposed symbolic solutions, reviewing the team's work and so on. A sense of these desirable group activities and the skill of designing problems to encourage them had to develop gradually through the design-based research iterations.

5.3.2 Non-laboratory Experimental Design (External Validity)

Of course, to isolate the small-group phenomena we do not literally isolate our subject groups from individuals and communities. The groups consist of students, who are individuals and who make individual contributions to the group discourse based on their individual readings of the discourse. In addition, the groups exist and operate within community and social contexts, drawing upon the language and practices of their math courses and of their teen and online subcultures. These are essential features of a real-world context and we would not wish to exclude them even to the extent possible by confining the interaction to a controlled laboratory setting. We want the students to feel that they are in a natural setting, interacting with peers. We do not try to restrict their use of language in any way (e.g., by providing standardized prompts for chat postings or scripting their interactions with each other).

We are designing a service that can be used by students and others under a broad array of scenarios: integrated with school class work, as extra-curricular activities, as social experiences for home-schooled students, as cross-national team adventures or simply as opportunities (in a largely math-phobic world) to discuss mathematics. To get a sense of how such activities might work, we have to explore interactions in naturalistic settings, where the students feel like they are engaged in such activities rather than being laboratory subjects.

5.3.3 Data Collection at the Group Level of Description (Unit of Analysis)

Take the network of references in a chat threading diagram as an image of meaning making at the group level (Stahl 2007). One could almost say that the figure consists entirely of contributions from individuals (the chat postings and white-board drawings) and resources from the math community; that everything exists on either the individual or community level, not on the group level. Yet, what is important in the figure is the network of densely interwoven references, more than the objects that are connected by them. This network exists at the group level. It mediates the individual and the community by forming the joint problem space (Sarmiento 2007; Teasley and Roschelle 1993), indexical ground (Hanks 1992), referential network (Heidegger 1927/1996) or situation (Suchman 2007) within which meanings, significant objects and temporal relations are intersubjectively co-constructed (Dourish 2001). On the individual level, these shared group meanings are interpreted and influence the articulation of subsequent postings and actions. On the community level, the meanings may contribute to a continually evolving culture through structuration processes (Giddens 1984). The VMT Project is oriented toward the processes at the group unit of analysis, which build upon, connect and mediate the individual and community phenomena.

Elements from the individual and community levels only affect the group level if they are referenced in the team's interaction. Therefore, we do not need to gather data about the students or their communities other than what appears in the interaction record. We do not engage in surveys or interviews of the students or their teachers. For one thing, the design of the VMT Project prohibits access to these sources of data, because the students are only available during the chat sessions. External sources of data would be of great interest for other research questions having to do with individual learning or cultural changes, but for our research question, they are unnecessary and might even form a distraction or skew our analysis because it would cause our readings of the postings to be influenced by information that the group had not had.

By moving to the disembodied online realm of group cognition in virtual math teams, it is easier for us to abandon the positivist metaphors of the mechanistic worldview. Not only is it clear that the virtual group does not exist in the form of a physical object with a persistent memory akin to a computer storage unit, but even the individual participants lack physical presence. All that exists when we observe the replayed chats are the traces of a discourse that took place years ago. Metaphors that might come naturally to an observer of live teamwork in a workplace or classroom—personalities, the group, learning, etc.—no longer seem fundamental. What exist immediately are the textual, graphical and symbolic inscriptions. These are significant fragments, whose meaning derives from the multi-layered references to each other and to the events, artifacts and agents of concern in the group discourse. This meaning is as fresh now as when the discourse originated, and can still be read off the traces by an analyst, much as by the original participants. This shows that the meanings shared by the groups are not dependent upon mental states of the individual students—although the students may have had interpretations of those meanings in mind, external to the shared experience. The form of our data reinforces our focus on the level of the shared-group-meaning making as an interactional phenomenon rather than a mental one.

5.3.4 Instrumentation and Data Formats (Objectivity)

When one videotapes small-group interactions a number of practical problems arise. Data on face-to-face classroom collaboration runs into issues of (a) recording and transcribing the verbal interaction, (b) capturing the visual interaction and (c) knowing about all the influences on the interaction. The data is in effect already partially interpreted by selective placement of the microphone and camera. It is further interpreted by transcription of the talk and is restricted by limited access to facial expressions and bodily gestures. Much happens in a classroom influencing the student teams that is not recorded.

The online setting of the VMT sessions eliminates many of these problems. As already described, the automatic computer log of the session captures everything that influences the group as a whole. This includes all the postings and whiteboard activity, along with their precise timing. They are captured at the same granularity as they are presented to the students. Chat postings appear as complete messages, defined by the author pressing the Enter button. Whiteboard textboxes appear as complete, when the author clicks outside of the textbox. Whiteboard graphics appear gradually, as each graphical element is positioned by the author. Computer-generated social-awareness messages (when people enter or exit the chat room, begin or end typing, move a graphical object, etc.) are also accurately recorded. The precision of the log recording is assured because it consists of the original actions (as implemented by the computer software) with their timestamps. The original display to the students is generated from the same data that is used by the VMT Replayer. There is no selectivity or interpretation imposed by the analysts in the preparation of the full session record.

For our analysis of chats, we use a VMT Replayer. The Replayer is simply an extended version of the Java applet that serves as the chat/whiteboard room in the VMT environment. The reproduced chat room is separated by a thin line at the bottom from a VCR-like interface for replaying the session (see Fig. 5.1). The session can be replayed in real time or at any integral multiple of this speed. It can be started and stopped at any point. An analyst can drag the pointer along the timeline to scroll both the whiteboard history and the chat history in coordination. One can also step through the recorded actions, including all the awareness messages. In addition, spreadsheet logs can be automatically generated in various useful formats.

The data analyzed in the VMT Project is recorded with complete objectivity. There is no selectivity involved in the data generation, recording or collecting process. Furthermore, the complete recording can be made available to other researchers as a basis for their reviews of our analyses or the conducting of their own analyses. For instance, there have been multiple published analyses of the VMT data by other research groups following somewhat different research questions, theories and methods. While collaborative sessions are each unique and in principle impossible to reproduce, it is quite possible to reproduce the unfolding of a given session from the persistent, comprehensive and replayable record (see Fig. 5.1).

5.3.5 Collaborative Data Sessions (Reliability)

Interpretation of data in the VMT Project first begins with an attempt to describe what is happening in a chat session. We usually start this process with a data session (Jordan and Henderson 1995) involving 6 to 12 researchers. A typical data session is

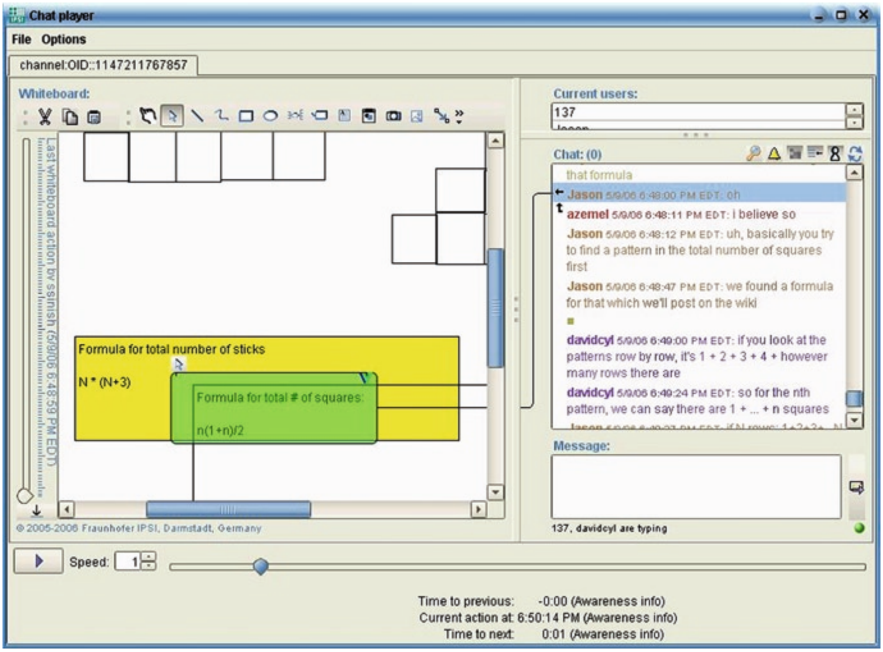

Fig. 5.1 The VMT Replayer

initiated by a researcher who is interested in having a particular segment of a session log discussed by the group. Generally, the segment seems to be both confusing and interesting in terms of a particular research question.

For our data sessions, we sit around a circle of tables and project an image of the VMT Replayer onto a screen visible to everyone. Most of us have laptop computers displaying the same Replayer, so that we can scan back and forth in the segment privately to explore details of the interaction that we may want to bring to the attention of the group. The group might start by playing the segment once or twice in real time to get a feel for how it unfolds. Then we typically go back to the beginning and discuss each line of the chat sequentially in some detail.

The interpretation of a given chat line becomes a deeply collaborative process. Generally, one person will make a first stab at proposing a hypothesis about the interactional work that line is doing in the logged discourse. Others will respond with suggested refinements or alternatives to the proposal. The group may then engage in exploration of the timing of chat posts, references back to previous postings or events, etc. Eventually the data analysis will move on to consider how the student group took up the posting. An interesting interpretation may require the analysts to return to earlier ground and revise their tentative previous understandings.

The boundaries of a segment must be considered as an important part of the analysis. When does the interaction of interest really get started and when is it resolved? Often, increasingly deep analysis drives the starting point back as we realize that earlier occurrences were relevant.

It is usually first necessary to clarify the referential structure of the chat postings and how they relate to events in the whiteboard or to the comings and goings of participants. The threading of the chat postings provides the primary structure of the online, text-based discourse in much the same way that turn taking provides the core structure of spoken informal conversation. Because of the overlap in the typing of chat postings, it is sometimes tricky to figure out who is responding to what. Looking at the timestamps of posts and even at the timestamps of awareness messages about who is typing can provide evidence about what was visible when a posting was being typed. This can often suggest that a given post could or could not have been responding to a specific other post, although this is sometimes impossible to determine. When it is hard for the analyst to know the threading, it may have also been hard for most of the chat participants (other than the typist) to know; this may result in signs of trouble or misunderstandings in the subsequent chat.

The test of *correctness* of chat interaction analysis is not a matter of what was in individuals' minds, but of how postings function in the interaction. Most of the multi-layered referencing takes place without conscious awareness by the participants, who are experts at semantic, syntactic and pragmatic referencing and can design utterances in response to local resources without formulating explicit plans (Suchman 2007). Thus, inspection of participants' memories would not reveal causes. Of course, participants could retroactively tell stories about why they posted what they did, but these stories would be based upon their current (not original) interpretations using their linguistic competence and upon their response to their current (not original) situation, including their sense of what the person interviewing them wants to hear. Thus, interpretations by the participants are not in principle privileged over those of the analyst and others with the relevant interpretive competence (Gadamer 1960/1988). The conscious memories that a participant may have of the interaction are, according to Vygotsky's theory, just more interaction—but this time sub-vocal self-talk; if they were brought into the analysis, they would be in need of interpretation just as much as the original discourse.

Since our research question involves the group as the unit of analysis, we do not raise questions in the data session about what one student or another may have been doing, thinking or feeling as an individual. Rather, we ask what a given posting is doing interactionally within the group process, how it responds to and takes up other posts and what opportunities it opens for future posts. We look at how a post is situated in the sequential structure of the group discourse, in the evolving social order and in the team's meaning making. What is this posting doing here and now in the referential network? Why is it "designed to be read" (Livingston 1995) in just this way? How else could it have been phrased and why would that not have achieved the same effect in the group discourse?

We also look at how a given posting *positions* (Harré and Moghaddam 2003) both the author and the readers in certain ways. We do not attribute constant personalities or fixed roles to the individuals, but rather look at how the group is organized through the details of the discourse. Perhaps directing a question toward another student will temporarily bestow upon her a form of *situated expertise*

(Zhou et al. 2008) such that she is expected to provide an extended sequence of *expository* postings (Mercer and Wegerif 1999).

The discussion during a data session can be quite unorderly. Different people see different possible understandings of the log and propose alternative analyses. Generally, discussion of a particular posting continues until a consensus is tentatively established or someone agrees to look into the matter further and come back next week with an analysis. Notes are often taken on the data session's findings, but the productive result of the discussion most often occurs when one researcher is inspired to write about it in a conference paper or dissertation section. When ideas are taken up this way, the author will usually bring the more developed analysis back for a subsequent data session and circulate the paper.

In coding analysis, it is conventional to train two people to code some of the same log units and to compare their results to produce an inter-rater reliability measure (Strijbos and Stahl 2007). In our chat interaction analysis, we do not pretend that the log can be unproblematically partitioned into distinct units, which can be uniquely assigned to a small number of unambiguous codes. Rather, most interesting group discourse segments have a complex network of interwoven references. The analysis of such log segments requires a sophisticated human understanding of semantics, interpersonal dynamics, mathematics, argumentation and so on. Much is ultimately ambiguous and can be comprehended in multiple ways—sometimes the chat participants were intentionally ambiguous. At the same time, it is quite possible for analysts to make mistakes and to propose analyses that can be shown to be in error. To ensure a reasonable level of reliability of our analyses, we make heavy use of data sessions. This ensures that a number of experienced researchers agree on the analyses that emerge from the data sessions. In addition, we try to provide logs—or even the entire session data with the Replayer—in our papers so that readers of our analyses can judge for themselves the interpretations that are necessarily part of chat analysis.

The collaborative analytic work of data sessions, as described above, is a central form of data analysis in the science of group cognition as practiced in the VMT Project. Coding of utterances, controlled experimental designs and statistical comparisons are not central to the methodology. However, attempts are often made to consider whether interaction patterns that emerge can be generalized as typical methods of interaction.

5.3.6 Describing Social Practices (Generalizability)

The research question that drives the VMT Project is: What are the distinctive mechanisms or processes that take place at the small-group level of description when the group is engaged in problem-solving or knowledge-building tasks? Therefore, we are interested in describing the inter-personal practices of the groups that interact in the VMT environment. There are, of course, many models and theories in the learning sciences describing the psychological practices of *individuals* involved in learning. At the opposite extreme, Lave and Wenger's (1991) theory of situated learning describes social practices of *communities* of practice, whereby a community renews itself by moving newcomers into increasingly central forms of legitimate peripheral participation. However, there are few descriptions specifically of how *small groups* engage in learning practices.

Vygotsky (1930/1978) argued that learning takes place inter-subjectively (in dyads or groups) before it takes place intra-subjectively (by individuals). For instance, in his analysis of the infant and mother (p. 56), he outlines the process through which an infant's unsuccessful grasping at some object becomes established by the mother-child dyad as a pointing at the object. This shared practice of pointing subsequently becomes ritualized by the dyad (LeBaron and Streeck 2000) and then mediated and "internalized" by the infant as a pointing gesture. The pointing gesture—as a foundational form of deictic reference—is a skill of the young child, which he can use for selecting objects in his world and learning about them. The gesture is understood by his mother because it was intersubjectively established with her. In this prototypical example, Vygotsky describes learning as an inter-subjective or small-group practice of a dyad.

While we can imagine that Vygotsky's description is based on a concrete interaction of a specific infant and mother in a particular time and place, the pointing gesture that he analyzed is ubiquitous in human culture. In this sense, the analysis of a unique interaction can provide a generalizable finding. The science of ethnomethodology (the study of the methods used by people) (Garfinkel 1967) is based on the fact that people in a given culture or linguistic community share a vast repertoire of social practices for accomplishing their mundane tasks. It is only because we share and understand this stock of practices that we can so quickly interpret each other's verbal and gestural actions, even in novel variations under unfamiliar circumstances. The analysis of unique case studies can result in the description of social practices that are generalizable (Maxwell 2004). The methods developed in specific situated encounters are likely to be typical of a broad range of cases under similar conditions.

In our data sessions, we find the same kinds of moves occurring in case after case that we analyze. On the one hand, group practices are extremely sensitive to changes in the environment, such as differences in features and affordances of the communication media. On the other hand, groups of people tend to adapt widespread methods of interaction to changing circumstances in similar ways—to support general human and social needs. Group practices are not arbitrary, but draw on rich cultural stocks of shared behavior and adapt the outward appearances in order to maintain the underlying structure under different conditions.

By describing the structure of group practices in detailed case studies, we can characterize general methods of group behavior, group learning or group cognition. Findings from analyses of case studies can lead to the proposal of theoretical categories, conceptualizations, structures or principles—in short, to a science of group interaction.

5.4 Conceptualizing Group Cognition

As discussed above, students in virtual math teams are active as individuals, as group participants and as community members. They each engage in their own, private *individual* activities, such as reading, interpreting, reflecting upon and typing chat messages. Their typed messages also function as *group* actions, contributing to the on-going problem solving of the team. Viewed as *community* events, the chats participate in the socialization process of the society, through which the students become increasingly skilled members of the community of mathematically literate citizens.

A thesis of the theory of group cognition is, "Small groups are the engines of knowledge building. The knowing that groups build up in manifold forms is what becomes internalized by their members as individual learning and externalized in their communities as certifiable knowledge" (Stahl 2006b, p. 16). Despite their centrality, small groups have not been theorized or studied extensively.

Some small-group literature has been produced from either the methodological perspective of psychology or that of sociology, primarily since World War II. Traumatized by the mass-culture horrors of fascism and by extreme forms of mentalist pseudo-science, these predominantly behaviorist studies focused on the negative aspects of "group think" and caricatured the notion of "group mind"—which had a well-respected history before the rise of positivism (Wegner 1986).

More recent theories like distributed cognition, situated action or activity theory actually conduct case studies of small-group interaction, but they do not theorize the small group as their unit of analysis and therefore they do not produce descriptions of small-group practices as such. Even Hutchins (1996), in studying distributed cognition in the wild, does not thematize the interpersonal interactions, but focuses on the cognitive unit of analysis, simply broadening it to include the external computational and physical representational artifacts that an individual worker uses. Furthermore, the cognitive accomplishments he studies are routine, well scripted procedures that do not involve creative solutions to ill-structured problems; the coordination of the navigational team is fixed by naval protocol, not co-constructed through the interaction.

The VMT studies provide a model for describing the small-group practices as distinct from individual and community processes. They look at rich interactions in groups larger than dyads, where individual identities play a smaller role. They analyze group efforts in high-order cognition such as mathematical problem solving and reflection on their problem-solving trajectory. They investigate groups that meet exclusively online, where the familiar visual, physical and aural modes of communication are unavailable, and where communication is mediated by designed technological environments. A number of findings are prominent in these analyses.

We shall review two findings here: One is that much group work is sustained and driven forward by *proposals* and responses to them. Another is that group interactions form a *social order*, which can often be characterized in terms of a temporal dimension, a joint problem space and an interaction space.

5.4.1 Proposal-Driven Sustained Group Activity

Careful review of many VMT logs shows that group interaction in these sessions is driven forward and sustained by various kinds of proposals. One of the first findings of the VMT Project was the role of "*math proposal adjacency pairs*" (Stahl 2006b, ch. 21 esp. pp. 442–456). These are simply a form of proposal adjacency pairs as found in informal face-to-face conversation, except that they deal with mathematical matters and they are only "adjacent" once their timing has been adjusted for threading. Technically, they might better be termed "math proposal response pairs," except that the term "adjacency pair" brings in the valuable theoretical connotations from conversation analysis (Sacks et al. 1974).

A proposal is not a solitary speech act. It involves minimally two acts (by two interacting people): a bid and a response. For instance, a question is only gradually formulated. People respond to an original opening bid and thereby define the question as an activity taken up in a certain way by the group (Zhou et al. 2008). Proposals generally, and math proposals more specifically, also have this structure:

- Someone posts a chat message or engages in some other activity that is designed to be read as a math proposal bid.
- This may begin to identify a math object as a potential focus of future group work.
- It is also designed to create possible responses, such as acceptances of a proposal for math work by the group.
- A second actor may respond to the bid as a proposal bid and accept it on behalf of the group, meaning that the group should work on it.
- The responder can alternatively reject the proposal on behalf of the group.
- The responder or additional group members can delay acceptance by posing a clarification question, for instance.
- Many other options and further steps are possible.

Through the proposal co-construction process, the group work becomes "object-oriented." The group orients to some mathematical object. Early in a session, the object may be based on a phrase from the task set for the group by the organizers of the VMT session. Later, it may be explicated by the group members in terms of visual representations or graphical objects in the whiteboard or symbolic math expressions in the chat. As group work continues through a series of many linked proposals, the math object to which the group orients may be a growing tree of multiple realizations of a math concept like *grid-world path*, *stair-step pattern*, *diamond pattern* or *hexagonal array*. The making of math proposals can be a mechanism for the *objectification* of a math object (Çakir et al. 2009).

The idea that group activity is strongly "object-oriented" is an important principle of activity theory (Engeström and Toiviainen 2009; Kaptelinin and Nardi 2006). It stresses the task-driven nature of group work. In the occupational settings that activity theory generally studies, activities often aim to accomplish a goal that has been established in advance (e.g., by management) as the purpose of the group. By highlighting the role of proposals as important means of structuring group interaction, the VMT studies of learning settings reveal a key interactional mechanism by means of which groups co-construct their own work goals in concrete detail.

Student groups in VMT sessions are highly responsive to the tasks that are pre-defined before they enter the chat room. These tasks are stated for them in various ways—on special web pages and/or by the moderator in chat—and the students clearly orient to them. However, one of the first things that the student group does is to discuss the task they will pursue. This is often put in the form of a posting like, "OK, let's figure out...." This is a proposal for what the group should work on next. It is selective of some feature of a broader task that was given to the group. As a proposal, it elicits a response from the rest of the group. The response further develops the proposed task. By highlighting the structure of the proposal, the analyses of the VMT Project show how the group itself accomplishes object orientation as an interactional achievement of the group. The object of a group's work is not given in advance and fixed for all time. Nor is it defined only at the level of a goal for the whole session. It is worked out and continually refined by the group interaction, even if it references texts and motivations from outside the group discourse. Furthermore, objects that orient the group work are proposed for small sequences of interaction as well as for the session-long sequences, as each new proposal is taken up.

The proposal structure introduces a temporal structure. A proposal often puts forward a task for the group to take on in the (near) future, possibly as a next step in its work. Sometimes—like at the end of a session that will be followed by another session of the same team—a proposal will plan for a future session. By its nature, a proposal bid creates possible next actions for the group, such as accepting, rejecting, questioning or ignoring the bid. In turn, the second part of the math proposal pair references back to the first part, which by now exists in the interaction past. It may well also reference events further back in the team's past, such as work already done or decisions previously made. The proposal as a whole, as it unfolds over potentially many actions, is always situated firmly in the present network of references. Thus, the proposal process contributes to establishing the temporal dimension of the group's work, with references to future, past and present events.

The proposals also serve to structure the temporal flow of the group interaction into episodes. They often define coherent sequences of discussion on the proposed topic, with openings and closings of the sequence. An episode of discussion on a given topic will typically be opened by a proposal bid, which begins to define the object of discussion. A protracted discussion may be closed by a new proposal that changes topic. Proposals operate on multiple scales: there may be a proposal about the object for a whole session, with proposals for large episodes of discussion within the session and proposals for detailed steps in the work. This provides a multi-layered temporal structure that can be analyzed at various granularities.

It is common to make diagrams of the proposal-response structure of chats. Such representations can be an important part of a science. The response structure, uptakes, adjacency pairs, sequences, etc. are central to an analysis of a chat interaction.

This theme is familiar in the broader literature on chat. The diversity of representations proposed (each with their rationale) indicates that this is a problematic issue as well as an important one for a future science of group cognition. Similarly, many researchers try to develop and apply coding schemes to analyze chats. A science of group cognition will have to take a stand on coding and on the appropriateness of specific coding schemes to interaction analysis (Strijbos and Stahl 2007).

The temptation to develop automated software (Erkens and Janssen 2008; Rosé et al. 2008) to construct graphical representations of the response structure and to categorize utterances may ironically serve to highlight the issues involved in making simplistic assumptions about the objective nature of the response structure and of the utterance character. A threading or uptake graph may make it look like postings exist with measurable attributes and fixed relationships, like the objects of Newtonian mechanics, with their precise location, mass and velocity. However, chat messages are more analogous to quantum particles, with their indeterministic and probabilistic characteristics. Whether a posting is a math proposal, a question or a joke depends on how an interpretive, thread-producing "reading" of it not only construes its uptake by subsequent postings, but also how it situates that posting in relation to previous postings. A particular posting may reference past and current artifacts, events and agents, but it also projects relevant "nexts," responses or uptakes by opening a field of possibilities. This is more complicated and less well defined than implied by a static diagram of nodes and links, however useful such a diagram may be to support visual reasoning about specific issues involving the flow of a chat. It may make more sense to treat postings as *mediating agents* in Latour's (2007) sense, as an alternative to metaphors from mechanistic theories of causation.

Proposal structures in VMT data can be more complicated than traditional analyses of adjacency pairs in studies of talk-in-interaction. Most case studies inspired by conversation analysis look at short sequences like a single adjacency pair or a pair that is temporarily interrupted by clarifications or repairs. The VMT Spring Fests allow analysis of longer sequences (Sarmiento 2007). In these, one sees mechanisms by means of which the work of a group is integrated into a layered temporal unity. The study of proposal mechanisms may lead to the identification of social structure in groups.

5.4.2 The Social Order of Group Cognition

Temporal structure is one dimension of the social order that a collaborative small group co-constructs of, by and for its interaction. Proposals are but one interactive mechanism for establishing the social order that supports the achievement of group cognition. By looking at bridging methods in longer sequences and across temporal and other discontinuities, analyses of VMT chats (Sarmiento-Klapper 2009) demonstrate the importance of the temporal dimension in addition to the content and relational dimensions that had been proposed by previous related research

(Barron 2000). This suggests three dimensions to the social order established by virtual math teams and other small groups engaged in group cognition:

- The *temporal dimension* of ordered events.
- The *problem space* of shared knowledge artifacts.
- The *interaction space* of positioned actors.

The first dimension of social order, the *temporal dimension*, was just discussed in terms of the ways in which proposal interactions are themselves temporally structured, with references to possible next responses, past resources and the current situation. The temporal dimension is also woven as part of the referential network of meaning that is built up through the group discourse. In particular, temporal indexicals (like *then*) and verb tenses establish the indexical ground of deictic reference (Hanks 1992), which is part of the shared meaning structure that makes sense of references to events and locates them within their temporal ordering.

In discourses about math, the second dimension, the *problem space*, is traditionally conceived of within the cognitivist tradition (Newell and Simon 1972) as a mental representation of mathematical relationships. The analysis of the work of virtual math teams (Çakir et al. 2009) shows that the group works out a shared notion of the math object, for instance by constructing visualizations in the whiteboard and instructing the group members to see them in a certain way. There is often a coordinated movement back and forth between visual, narrative and symbolic reasoning that gradually objectifies the math object into a rich, interconnected, meaningful multiplicity of significances and realizations. The representation of the object for the group does not lie hidden in individual minds like the data structure of an artificial intelligence software system. It consists of a network of visible inscriptions in the visual interface of the VMT environment, tied together into a meaningful whole by the set of carefully crafted references within the group interaction. The object exists as an *artifact*, a physical object that is meaningful (Stahl 2006b, ch. 16). However, in the case of math objects that are the result of extensive group work, there is not a single identifiable artifact; the math object consists of a "tree of multiple realizations" (Sfard 2008; Stahl 2008a) united by the group discourse and only imperfectly objectified in a single phrase or symbol.

In particular, once the rich experience of the group interaction that built the math artifact is summarized or sedimented into a single sign and passed on to others who were not involved in the original experience (e.g., late-comers or newcomers), the full meaning of the artifact is hard to come by. This is the problem of math education. For new individuals to build anything like a mental representation of a math artifact, they need to go through a process like that which Vygotsky termed *internalization*. Either they need to experience a group process like those that occur in virtual math teams or they need to simulate such a process on their own. One often sees math students sketching visual reasoning diagrams on paper, playing around with symbolisms and arguing with themselves as though they were acting out the parts of a complete team. The path to math comprehension seems to require the practices of group problem solving, which experienced experts have learned to individuate and to conduct as individuals, imagining the visualizations and speaking the discourse sub-vocally.

The third dimension of social order is the *interaction space* of intersubjective relations. We characterize this in terms of *positioning* (Harré and Moghaddam 2003). In the VMT environment, there is no power hierarchy or other system of roles among the students. (The adult mentor who may be in the chat room with the students is, of course, an authority figure, but tends to play a minimal role in the session and rarely enters into the math work or interactions among the team. The mentor is positioned as being outside of the team, often by the mentor's own postings.) Researchers often discuss collaboration in terms of roles (Strijbos et al. 2004). They even advocate scripting or assigning fixed roles to students to make sure that certain functions of group process are carried out—such as leading the discussion, watching the time allotted for the session, summarizing the group accomplishments, monitoring the active participation of all members, controlling turn taking. In contrast to such an imposed approach, an analysis in terms of positioning views roles as fluidly changing, based on details of the group discourse.

Perhaps the clearest example of positioning arises in questioning. When one person asks another what some term means or how a result was derived, the questioner may be positioned as lacking knowledge and the addressee as having *situated expertise*. What this means is that the first person cedes the second the floor. The questioner will refrain from posting anything for a while and will expect the other group members to do likewise while the second person—the temporary expert for purposes of this question—will be expected to post a series of expository messages responding to the question. Questions are carefully designed to engage in positioning moves and other interpersonal work. Through methods like questioning and displays of individual knowledge, group members co-construct the intersubjective fabric of the group, often starting from a condition where there are no differentiations.

5.5 How We Study Group Cognition

A science of group cognition aims to comprehend the group processes that accomplish cognitive tasks like problem solving in collaborative math discourse. We have just reviewed quickly what has already been discovered in the VMT Project about how groups sustain their inquiries with various methods of making proposals. We have briefly characterized the social order of group cognition as involving dimensions of temporality, a problem space and an interaction space. These preliminary results were produced through dozens of case studies of excerpts from VMT logs. We have argued that the data analysis was conducted through scientific inquiry that can be valid, objective, reliable and generalizable in pursuing group cognition research questions at the group unit of analysis.

References

Adams, F., & Aizawa, K. (2008). *The bounds of cognition*. Malden: Blackwell.

Barron, B. (2000). Achieving coordination in collaborative problem-solving groups. *Journal of the Learning Sciences, 9*(4), 403–436.

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

Cakır, M. P., Zemel, A., & Stahl, G. (2009). The joint organization of interaction within a multimodal CSCL medium. *International Journal of Computer-Supported Collaborative Learning, 4*(2), 115–149. Available at http://dx.doi.org/10.1007/s11412-009-9061-0.

Çakir, M. P., Zemel, A., & Stahl, G. (2009). The joint organization of interaction within a multimodal CSCL medium. *International Journal of Computer-Supported Collaborative Learning (ijCSCL), 4*(2), 115–149.

Charles, E. S., & Shumar, W. (2009). Student and team agency in VMT. In G. Stahl (Ed.), *Studying virtual math teams* (pp. 207–224). New York: Springer. Available at http://dx.doi.org/10.1007/978-1-4419-0228-3_11.

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

Cui, Y., Kumar, R., Chaudhuri, S., Gweon, G., & Rosé, C. P. (2009). Helping agents in VMT. In G. Stahl (Ed.), *Studying virtual math teams* (pp. 335–354). New York: Springer. Available at http://dx.doi.org/10.1007/978-1-4419-0228-3_19.

Dourish, P. (2001). *Where the action is: The foundations of embodied interaction*. Cambridge: MIT Press.

Engeström, Y. (1987). *Learning by expanding: An activity-theoretical approach to developmental research*. Helsinki: Orienta-Kosultit Oy.

Engeström, Y. (2008). *From teams to knots*. Cambridge: Cambridge University Press.

Engeström, Y., & Toiviainen, H. (2009). Co-configurational design of learning instrumentalities: An activity-theoretical perspective. In S. Ludvigsen, A. Lund, & R. Säljö (Eds.), *Learning in social practices. ICT and new artifacts: Transformation of social and cultural practices*. Pergamon.

Erkens, G., & Janssen, J. (2008). Automatic coding of dialogue acts in collaboration protocols. *International Journal of Computer-Supported Collaborative Learning, 3*(4), 447–470. Available at http://dx.doi.org/10.1007/s11412-008-9052-6.

Gadamer, H.-G. (1960/1988). *Truth and method*. New York: Crossroads.

Garfinkel, H. (1967). *Studies in ethnomethodology*. Englewood Cliffs: Prentice-Hall.

Giddens, A. (1984). Elements of the theory of structuration. In *The constitution of society* (pp. 1–40). New York: University of California Press.

Hanks, W. (1992). The indexical ground of deictic reference. In A. Duranti & C. Goodwin (Eds.), *Rethinking context: Language as an interactive phenomenon* (pp. 43–76). Cambridge: Cambridge University Press.

Harré, R., & Moghaddam, F. (2003). Introduction: The self and others in traditional psychology and in positioning theory. In R. Harré & F. Moghaddam (Eds.), *The self and others* (pp. 1–11). Westport: Praeger.

Heidegger, M. (1927/1996). *Being and time: A translation of Sein und Zeit (J. Stambaugh, Trans.)*. Albany: SUNY Press.

Hutchins, E. (1996). *Cognition in the wild*. Cambridge: MIT Press.

Jordan, B., & Henderson, A. (1995). Interaction analysis: Foundations and practice. *Journal of the Learning Sciences, 4*(1), 39–103. Available at http://lrs.ed.uiuc.edu/students/c-merkel/document4.HTM.

Kaptelinin, V., & Nardi, B. A. (2006). *Acting with technology: Activity theory and interaction design*. Cambridge: MIT Press.

Koschmann, T. (1996). Paradigm shifts and instructional technology. In T. Koschmann (Ed.), *CSCL: Theory and practice of an emerging paradigm* (pp. 1–23). Mahwah: Lawrence Erlbaum.

Latour, B. (2007). *Reassembling the social: An introduction to actor-network-theory*. Cambridge: Cambridge University Press.

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

LeBaron, C., & Streeck, J. (2000). Gesture, knowledge and the world. In D. McNeill (Ed.), *Language and gesture* (pp. 118–138). Cambridge: Cambridge University Press.

Livingston, E. (1995). *An anthropology of reading*. Bloomington: Indiana University Press.

Maxwell, J. (2004). Causal explanation, qualitative research, and scientific inquiry in education. *Educational Researcher, 33*(2), 3–11. Available at http://www.aera.net/pubs/er/pdf/vol33_02/2026-02_pp03-11.pdf.

Medina, R., Suthers, D. D., & Vatrapu, R. (2009). Representational practices in VMT. In G. Stahl (Ed.), *Studying virtual math teams* (pp. 185–205). New York: Springer. Available at http://dx.doi.org/10.1007/978-1-4419-0228-3_10.

Mercer, N., & Wegerif, R. (1999). Is "Exploratory talk" Productive talk? In K. Littleton & P. Light (Eds.), *Learning with computers: Analyzing productive interaction* (pp. 79–101). New York: Routledge.

Mühlpfordt, M., & Stahl, G. (2007). The integration of synchronous communication across dual interaction spaces. In C. Chinn, G. Erkens, & S. Puntambekar (Eds.), *The proceedings of CSCL 2007: Of mice, minds, and society (CSCL 2007)*. New Brunswick, NJ. Available at http://GerryStahl.net/vmtwiki/martin.pdf.

Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs: Prentice-Hall.

Piaget, J. (1985). *The equilibrium of cognitive structures: The central problem of intellectual development*. Chicago: Chicago University Press.

Powell, A. B., & Lai, F. F. (2009). Inscriptions, mathematical ideas and reasoning in VMT. In G. Stahl (Ed.), *Studying virtual math teams* (pp. 237–259). New York: Springer. Available at http://dx.doi.org/10.1007/978-1-4419-0228-3_13.

Robbins, P., & Aydede, M. (Eds.). (2009). *The Cambridge handbook of situated cognition*. Cambridge: Cambridge University Press.

Rosé, C., Wang, Y.-C., Cui, Y., Arguello, J., Stegmann, K., Weinberger, A., et al. (2008). Analyzing collaborative learning processes automatically: Exploiting the advances of computational linguistics in computer-supported collaborative learning. *International Journal of Computer-Supported Collaborative Learning, 3*(3), 237–271. Available at http://dx.doi.org/10.1007/s11412-007-9034-0.

Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language, 50*(4), 696–735. Available at www.jstor.org.

Sarmiento, J. (2007). Bridging: Interactional mechanisms used by online groups to sustain knowledge building over time. Paper presented at the international conference on Computer-Supported Collaborative Learning (CSCL '07), New Brunswick, NJ., Available at http://GerryStahl.net/vmtwiki/johann.pdf.

Sarmiento, J., & Stahl, G. (2008). Extending the joint problem space: Time and sequence as essential features of knowledge building. Paper presented at the International Conference of the Learning Sciences (ICLS 2008), Utrecht, Netherlands. Available at http://GerryStahl.net/pub/icls2008johann.pdf.

Sarmiento-Klapper, J. W. (2009). Bridging mechanisms in team-based online problem solving: Continuity in building collaborative knowledge. Unpublished Dissertation, Ph.D., College of Information Science and Technology, Drexel University, Philadelphia, PA.

Sawyer, R. K. (2003). *Group creativity: Music, theater, collaboration*. Mahwah: Lawrence Erlbaum.

Schwartz, D. (1995). The emergence of abstract representations in dyad problem solving. *Journal of the Learning Sciences, 4*(3), 321–354.

Sfard, A. (2008). *Thinking as communicating: Human development, the growth of discourses and mathematizing*. Cambridge: Cambridge University Press.

Stahl, G. (2002). Groupware goes to school. In J. H. J. Pino (Ed.), *Groupware: Design, implementation and use: Proceedings of the 8th international workshop on groupware (CRIWG '02)* (Vol. 2440, pp. 7–24) La Serena, Chile. Springer. Available at http://GerryStahl.net/cscl/papers/ch11.pdf.

Stahl, G. (2004a). Can community knowledge exceed its members'? *ACM SigGroup Bulletin, 23*(3), 1–13. Available at http://GerryStahl.net/publications/conferences/2003/group/workshop.pdf.
Stahl, G. (2004b). Mediation of group cognition. *SigGroup Bulletin, 24*(4), 13–17. Available at http://GerryStahl.net/publications/journals/Mediation%20of%20Group%20Cognition.pdf.
Stahl, G. (2005). Group cognition: The collaborative locus of agency in CSCL. Paper presented at the international conference on Computer Support for Collaborative Learning (CSCL '05), Taipei, Taiwan (pp. 632–640). Lawrence Erlbaum Associates. Available at http://GerryStahl.net/pub/cscl2005.pdf & http://GerryStahl.net/pub/cscl2005ppt.pdf.
Stahl, G. (2006a). Analyzing and designing the group cognitive experience. *International Journal of Cooperative Information Systems (IJCIS), 15*, 157–178. Available at http://GerryStahl.net/pub/ijcis.pdf.
Stahl, G. (2006b). *Group cognition: Computer support for building collaborative knowledge*. Cambridge: MIT Press. Available at http://GerryStahl.net/mit/.
Stahl, G. (2006c). Supporting group cognition in an online math community: A cognitive tool for small-group referencing in text chat. *Journal of Educational Computing Research (JECR) special issue on Cognitive tools for collaborative communities, 35*(2), 103–122. Available at http://GerryStahl.net/pub/jecr.pdf.
Stahl, G. (2006d). Sustaining group cognition in a math chat environment. *Research and Practice in Technology Enhanced Learning (RPTEL), 1*(2), 85–113. Available at http://GerryStahl.net/pub/rptel.pdf.
Stahl, G. (2007). Meaning making in CSCL: Conditions and preconditions for cognitive processes by groups. Paper presented at the international conference on Computer-Supported Collaborative Learning (CSCL '07), New Brunswick, NJ: ISLS, Available at http://GerryStahl.net/pub/cscl07.pdf.
Stahl, G. (2008a). Book review: Exploring thinking as communicating in CSCL. *International Journal of Computer-Supported Collaborative Learning, 3*(3), 361–368. Available at http://dx.doi.org/10.1007/s11412-008-9046-4.
Stahl, G. (2008b). Human-human interaction and group learning. Paper presented at the Human-Computer Interaction Consortium, Frasier, CO, Available at http://GerryStahl.net/pub/hcic2008.pdf.
Stahl, G. (2009a). From individual representations to group cognition. In G. Stahl (Ed.), *Studying virtual math teams* (pp. 57–73). New York: Springer. Available at http://dx.doi.org/10.1007/978-1-4419-0228-3_5.
Stahl, G. (2009b). *Studying virtual math teams*. New York: Springer. Available at http://GerryStahl.net/vmt/book.
Stahl, G., & Herrmann, T. (1999). Intertwining perspectives and negotiation. Paper presented at the ACM SIGGROUP Conference on Supporting Group Work (Group '99), Phoenix, AZ (pp. 316–324). Available at http://GerryStahl.net/cscl/papers/ch07.pdf.
Strijbos, J. W., Martens, R. L., Jochems, W. M. G., & Broers, N. J. (2004). The effect of functional roles on perceived group efficiency: Using multilevel modeling and content analysis to investigate computer-supported collaboration in small groups. *Small Group Research, 35*, 195–229.
Strijbos, J. W., & Stahl, G. (2007). Methodological issues in developing a multi-dimensional coding procedure for small group chat communication. Learning & Instruction. *Special issue on measurement challenges in collaborative learning research, 17*(4), 394–404. Available at http://GerryStahl.net/vmtwiki/jw.pdf.
Suchman, L. A. (2007). *Human-machine reconfigurations: Plans and situated actions* (2nd ed.). Cambridge: Cambridge University Press.
Teasley, S. D., & Roschelle, J. (1993). Constructing a joint problem space: The computer as a tool for sharing knowledge. In S. P. Lajoie & S. J. Derry (Eds.), *Computers as cognitive tools* (pp. 229–258). Mahwah: Lawrence Erlbaum Associates, Inc.
Trausan-Matu, S., & Rebedea, T. (2009). Polyphonic inter-animation of voices in VMT. In G. Stahl (Ed.), *Studying virtual math teams* (pp. 451–473). New York: Springer. Available at http://dx.doi.org/10.1007/978-1-4419-0228-3_24.
Vygotsky, L. (1930/1978). *Mind in society*. Cambridge: Harvard University Press.

Wee, J. D., & Looi, C.-K. (2009). A model for analyzing math knowledge building in VMT. In G. Stahl (Ed.), *Studying virtual math teams* (pp. 475–497). New York: Springer. Available at http://dx.doi.org/10.1007/978-1-4419-0228-3_25.

Wegner, D. (1986). Transactive memory: A contemporary analysis of the group mind. In B. Mullen & G. R. Goethals (Eds.), *Theories of group behavior* (pp. 185–208). New York: Springer.

Zhou, N. (2009). Question co-construction in VMT chats. In G. Stahl (Ed.), *Studying virtual math teams* (pp. 141–159). New York: Springer. Available at http://dx.doi.org/10.1007/978-1-4419-0228-3_8.

Zhou, N., Zemel, A., & Stahl, G. (2008). Questioning and responding in online small groups engaged in collaborative math problem solving. Paper presented at the International Conference of the Learning Sciences (ICLS 2008), Utrecht, Netherlands. Available at http://GerryStahl.net/pub/icls2008nan.pdf.

Chapter 6
Thinking about Methods to Capture Effective Collaborations

Stephanie D. Teasley

> *Peers do not learn because they are two, but because they perform some activities which trigger specific learning mechanisms*
>
> Dillenbourg (1999, p. 5).

The overall goal of this volume is to deepen our understanding of the ways to document and analyze interactions in CSCL and the first five chapters provide a variety of methodologies for doing so and the first five chapters provide a variety of methodologies for doing so. Specifically, these chapters provide different methods for operationalizing and analyzing the *some activities* referred to by Dillenbourg in the quote above. Each chapter provides a different perspective on an analytic problem that has vexed researchers of collaboration since the earliest work: characterizing how learning happens when people work together.

Theoretically, these chapters share the view that learning is a process whereby the social and cognitive are fundamentally intertwined. Although this was a somewhat more radical proposition in the earlier days of research on collaboration when development psychologists argued about "Are two heads better than one?" (e.g., Perret-Clermont 1980), today most learning scientists accept this viewpoint as a given. As a result, the focus of analysis has moved from individual outcomes to the social plane, where emergent conceptions are made visible and can be analyzed as a group product. Even with agreement about the fundamentally social basis for learning, the question about how to best conceptualize collaboration remains open. These chapters provide readers with compelling examples of different empirical methods to characterize the emergent and socially constructed properties of the collaborative process.

There is also a consensus among the authors of these chapters that collaboration can be beneficial to learning but the conditions under which it is successful are elusive and positive outcomes are not guaranteed. When the main research question

S.D. Teasley (✉)
School of Information, University of Michigan, 4384 North Quad,
105 S. State St., Ann Arbor, MI 48109–1285
e-mail: steasley@umich.edu

S. Puntambekar et al. (eds.), *Analyzing Interactions in CSCL*, Computer-Supported Collaborative Learning Series, DOI 10.1007/978-1-4419-7710-6_6,

for studies of collaboration turned to "*When* are two heads better than one?" the conditions for successful collaboration became the focus (see Dillenbourg et al. 1996). Despite a growing body of literature demonstrating which particular variables do or do not improve learning outcomes, the results of these kinds of studies still could not point clearly to *why* two heads can be better than one. The chapters presented here illustrate this shift in focus away from the products of collaborative work to an analysis of the process of collaboration. Each author has a somewhat different global term for describing "good collaboration:" meaning making (Arvaja), knowledge building (Law et al., Stahl), convergence (Kapur), and shared understanding (Hmelo-Silver et al.) and the level of analysis varies from the dyad to the classroom. Although there are some differences in how these authors conceptualize effective collaboration and the context in which they focus their research, each chapter provides a methodology for revealing how learning takes place.

6.1 Capturing the Influence of Context

The chapter by Arvaja provides data from two of her previous studies to illustrate how "socio-cultural discourse analysis" can be used to understand the contextual nature of collaboration. Here the emphasis is on using "a dialogical approach and discourse analysis" Arvaja, this volume to provide both conceptual and analytic tools for understanding how collaboration proceeds. This methodology relies heavily on the work of Gee and Green (1998) and Linell (1998) to reveal how consideration of three different general contexts of activity, the immediate (perceptual) context, the socio-cultural context, and the local context, are necessary to understand the embedded nature of collaborative activity.

In the first study Arvaja utilizes selected data from her 2008 ICLS conference paper examining two university students as they work on a course project to design a questionnaire for evaluating a web-based learning environment. The excerpts of the students' discussion demonstrate not only the dynamic nature of the their discourse, but also how their talk is shaped by the implicit rules of the cultural context (how to do research), prior knowledge (about questionnaire methodology), and identity (as students who know less that their professor). Through the use of these examples, the author makes clear the embedded nature of the students' discourse and what would be missed- or misunderstood- without consideration of the context of the activity.

The second study provides a methodology for closely analyzing the process of shared knowledge construction within the contextual framework illustrated in Study 1. Arvaja draws on data from her 2007 study of small groups of teacher education students collaborating via an asynchronous discussion forum. Here she demonstrates how to evaluate the students' shared knowledge construction using thematic content analysis to create a multidimensional coding scheme. While the "code and count" method has been widely utilized (and sometimes criticized) in the literature on collaboration, Arvaja's goal is a "deeper understanding" of the students' activity that she believes can come from (1) linking the coding with (2) a qualitative description and interpretation enabled by the three dimensions on analysis framework she presented in study one.

There are several lessons that can be taken away from Arvaja's chapter. The first is the structural format of the coding scheme. She justifies her decision process about the unit of analysis, the thematic meaning unit, and ties it closely to the intent of the coding scheme. This is a crucial decision point for analyzing discourse and the researcher needs to have a clear justification for making this choice, as has Arvaja. The larger coding categories, shown in Table 1 in chapter 2, are not strictly limited to this particular data set and could be used by other analysts. Perhaps more important than the coding particulars is that Arvaja's chapter demonstrates how one can utilize the by now fairly routine method of content coding (see the Hmelo-Silver et al. chapter) and increase its analytic value by closely examining how activity is deeply embedded in the participants' perspective, including their shared history, the task context, as well as their here-and–now experience during the activity. This is a method that relies on the talent of the analyst to properly identify and interpret the themes underlying the discourse, and requires rigorous standards for assessing whether the analyst is making valid assumptions about what is going on in the collaboration.

As Arvaja herself recommends, "additional data sources might be needed to fully ensure that the analyst has properly characterized the meaningful elements of the participants' own experience". For example, she recommends interviews or diary entries to "shed more light on (other) resources relevant from the students' perspective." What she doesn't suggest, however, is using other forms of data that are apparent to the analyst: participants' nonverbal behaviors. This falls more in the realm of interaction analysis (Jordan and Henderson 1995) than discourse analysis (Gee 2005), but nonetheless provides the analyst with the same resources that the participants themselves are using to form their own interpretation of events. For example, a student might report on a post-task survey that they felt a high level of task engagement, but behaviors such as leaning way back in their seat and staring out the window would suggest to the analyst (and the collaborators) that this student wasn't much interested in the task at hand. Having information about the students' shared history, as Arvaja recommends, could also then help account for the collaborators' reaction to such behavior. For example, the collaborators might ignore it if this student was known to be a free rider or try to re-engage the student if they were usually a more reliable colleague.

6.2 Representing Timing

The chapter by Hmelo-Silver and her colleagues is the first of three to include methods that focus on temporal features of the collaborative process. These chapters take head-on the question of what is missing when the collaboration is represented only by aggregated frequency counts of coded behaviors. In this chapter, Hmelo-Silver et al. provide a technique for visually representing multiple sources of data over time allowing an analysis of data generated from participants' face-to-face conversations as well as activity that occurs as participants interact with an online learning environment. The authors' intent here is to provide representations of collaboration that reveal the sequential relationship between discourse acts and how the talk relates to the tools available in the learning environment. The authors build on work from a number of researchers who have developed analytic tools for representing collaboration, primarily Luckin et al.'s (2001) CORDFU, to create a system to generate "CORDTRA diagrams" (Chronologically-Ordered Representation of Discourse and Tool-Related activity). The early part of the chapter provides a brief, but thorough overview of key research on visual representations for CSCL, which is a great starting point for readers new to this kind of analysis.

CORDTRA diagrams present a linear representation of activity plotted sequentially over time, with multiple processes comprising each line of data. This method allows the researcher to visually capture a variety of codes, including discourse, nonverbal activity such drawing or pointing, and tool-specific activity generated from the system log such as entries on a white board or discussion board. By creating visualizations of collaboration that capture multiple codes and display them in the order in which they occur, the authors provide an analytic tool that reveals how the activity unfolds and displays this process in the context of how participants use the resources available in learning system. The horizontal axis shows coded events identified by the analyst and the chronology of events is represented along the horizontal axis of the diagram. This makes patterns in the occurrence of individual events visible as well as the co-occurrence of multiple variables that cluster together at specific points in time. Production of the CORDTRA diagrams themselves is not the end goal of the method, but rather they provide a resource that allows the analyst to cycle between various levels of analysis.

Hmelo-Silver and her co-authors demonstrate the utility of the CORDTRA diagrams by providing representations drawn from their previous studies of collaborative learning in two kinds of computer-supported contexts: asynchronous interaction and synchronous interaction. In Example 1, they show how the CORDTRA technique can be used to look at the coding of students' online discourse together with log data from the system. By contrasting two groups, one successful and one unsuccessful, the diagrams shown in Figures 2–5 in chapter 4 reveal significant differences in students' coded discourse and system use that point to several characteristics of good collaboration. These differences include the frequency and timing of (1) facilitator (the teaching staff) input, (2) use of system resources, and (3) group monitoring. While the observed differences in frequency would have been clear from a "code and count" method, the additional information about when certain behaviors happened helped the researchers to better understand why one group was less successful than the other. For example, while both groups engaged in group monitoring, the successful collaborators did so throughout the activity and closely linked to their own discussions about the task. By contrast, the less successful group only did this kind of monitoring late in their work and mostly generated by the facilitator.

In Example 2, the authors present selected data from Lui (2008) to demonstrate how the CORDTRA technique can be used to look at students' face-to-face discourse together with log data from the system. The authors have already established that there are learning gains when middle school students work with a simulation of a complex biological system (RepTools), but here they use the CORDTRA diagrams to better understand why the students did so. Again the authors provide a contrast between two groups who had very different levels of success with the task. In this case, the additional information representing the sequence of discourse types (Figures 7 and 8 in chapter 4) helped to reveal the relationship between what the students talked about and showed different patterns for when these discourse types appeared. For example, the contrast between the two diagrams indicates that although both groups were generating a similar number of questions requiring explanations, the less successful group was more focused on the behavioral level of the data and they were constantly manipulating the simulation without moving between the levels of reasoning shown by the successful groups (and more typical of domain experts).

The CORDTRA technique presented in this chapter provides a tantalizing but also somewhat overwhelming picture of discourse and activity. The method is tantalizing because it suggests how the timing of talk and action can be used to reveal patterns of behavior that underlie the important aspects of effective collaborations. In addition, this technique is completely agnostic about the coding scheme used. The analyst can develop categories that are best suited to the particulars of the task, context, and resources available to the learners. Application of the coding requires the same care and standard for reliability needed for analyses generating aggregated frequency data. Insights offered in the Arvaja chapter can be integrated into this method as well.

Representing the coding, however, is much more labor intensive with this method and it is obviously sensitive to getting the timing right. With online discourse, time stamps provided in the system logs can help address this challenge. When the discourse is unmediated by technology, however, the quality of the transcription is critical and representation of the talk needs to be very deliberately synchronized with the activity data. Fortunately, the authors suggest that there are plans underway to automate the construction of the CORDTRA diagrams. But even when the diagrams are constructed automatically, the challenge remains for the analyst to make sense of these very complicated representations. Indeed, future versions of the system may need a visualizer for the visualizations (!), including tools allowing researchers to select and hide various aspects of the data to focus on selected variables and move between coding levels. Here Hmelo-Silver and her colleagues have deliberately selected contrasting cases where the differences between groups are, in the authors' experience, quite apparent. Yet there's quite a bit of cognitive effort needed to isolate the points of interest in the diagram, even when guided by the authors' text. Researchers inexperienced with this technique would likely require significant training before being able to utilize it effectively. The value in doing so, however, is well demonstrated by this chapter.

6.3 Calculating When Timing Matters

The chapter by Kapur and his colleagues presents a quantitative methodology for representing the temporal evolution of collaborative problem solving. Like Hmelo-Silver et al., these authors posit that timing matters, but what they add here is data showing that specific timing matters most. Kapur et al. provide a technique for analyzing interaction by applying a complex system perspective to discourse data. They use this method to tackle the question of how groups come to a shared understanding of a problem or reach "cognitive convergence." This term has achieved a fair level of popularity for characterizing the key processes in collaboration (Teasley et al. 2008) and Kapur's approach for understanding this phenomenon is unique in the literature on collaboration.

The major focus of this chapter is to show that cognitive convergence can be characterized mathematically, where each discourse move in an interaction can be coded according to its relationship to the group's goal. Using a scale of +1 to –1, each "interaction unit" in a turn can be coded for whether or not it moves the group closer to a correct problem solution (see Example in chapter 1). This coding then allows the progression of talk to be operationalized as a Markov walk (Ross 1996) which illuminates the evolution of convergence over time. Applying this technique to a data set consisting of online chat discussions between student triads working on both a well-structured and ill-structured problem showed that ultimate group performance can be predicted based on what happens in the first 30–40% of the talk.

There are two aspects of the technique demonstrated in this chapter that provide valuable lessons for other analysts of collaboration. The first is the representation of ongoing talk as a fitness curve (see Figure 1 in chapter 1). This necessitates that the analyst be able to accurately *and meaningfully* assign an impact value between +1 and –1 to each unit of talk. Kapur et al. have successfully done so with well- and ill-structured problems using a coding rubric to quantify the quality of the solution generated by the group (see Table 1 in chapter 1). When a group outcome (e.g., problem solution) can be coded as more or less successful, the resulting graphs provide clear pictures of the differences in the progress of discussion between high and low performing groups. These graphs also demonstrate nicely that high performing groups do not simply engage in a succession of convergent ideas and that there isn't only one canonical path to success.

The second lesson that can be taken away from this chapter is the importance of what happens in groups at the earliest stage of their working together. In many respects, this insight is not surprising as it is consistent with many of the classic findings in decision theory and social psychology. Concepts such as primacy effects (Nisbett and Ross 1980), anchoring effects (Tversky and Kahneman 1974) and first advocacy (Weisband 1992) have emerged from the empirical study of group behavior and demonstrate that what gets said first is a strong predictor of final outcomes. What Kapur et al. have provided here is a method by which analysts can determine how these effects play out in the context of collaborative learning.

The data from Kapur and his colleagues provide a strong challenge to the assumption of temporal homogeneity that underlies methods using frequency counts of specific behaviors. Both Kapur and Hmelo-Silver have demonstrated that it is very difficult to understand why a particular outcome happens when all you know is that a total number of behaviors X, Y, Z occurred. What remains open, however, is an understanding of the exceptions to Kapur's 30–40% rule: when an idea is introduced later in the problem solving process that sets the subsequent discussion on a better path. What are the conditions under which a novel and perhaps contradictory set of ideas is introduced into a group discussion and when/why is it not disregarded? Are mediated discussions more or less sensitive to timing effects? An integration of Hmelo-Silver's approach with Kapur's might provide a productive method for unpacking the particulars of why certain discourse moves, singly or in sequence, do or do matter at any given point during the collaborative process. Meanwhile we can use Kapur and his colleagues' results to think about the implications for scaffolding collaboration, focusing efforts early in the group work as suggested in the discussion section of this chapter.

6.4 Capturing Community Knowledge Building

The chapter by Law and her colleagues provides a methodology for looking at collaboration at a much larger scale than demonstrated in the previous chapters. Rather than focusing on dyadic or small group processes, these authors examine knowledge building at the community level. In the research reported here, they do so by looking at online discussions between school children in Hong Kong participating in the Knowledge Forum®. Knowledge Forum is a specialized form of a learning management system that supports a shared discourse environment with a set of scaffolds designed to support knowledge building (see Scardamalia, and Bereiter 2003). As such, logs of system use capturing individual contributions to a discussion (called "notes") provide a rich corpus of data that can be used to investigate how discussions involving large groups of learners can advance the overall collective knowledge available within a community.

Law et al. investigate the process of community knowledge building through the use of several tools that leverage machine coding to automate coding and analysis. The first is the Analytic Toolkit (Burtis 1998) that produces summary statistics of individual participants' activity in Knowledge Forum, including the number of notes created and viewed, and use of the system's support tools. The second is the Visual Intelligent Content Analyzer (Law et al. 2007) that matches specified text patterns and computes the semantic closeness of a string of discourse codes. The third is the Knowledge Space Visualizer (Teplovs and Scardamalia 2007) that displays graphically the relationships and interactions between discourse "notes." The results of an earlier study (Law et. al. 2008) showed no relationships between various measures captured by these tools and learning outcomes when the analyses were conducted at the individual student level. However, analyses conducted at the level of discourse thread, thus calculated on contributions created over time by multiple students, showed identifiable patterns in structure and scaffold use that distinguished productive discussion threads, especially over a sequence lasting at least five notes.

Based on the findings from the 2008 study, Law et al. examined seven additional discussion data sets from several classrooms using Knowledge Forum (see Table 1 in chapter 5). Here the authors set up coding schemes to capture several aspects of the students' behavior: use of scaffolds within the system, argumentation and questioning speech acts, and subject-related content as represented by keywords and key-word patterns. These codes were captured using the tools listed above and the data were represented chronologically using a linear visualizer called the Bobinette tool. Results from this method are presented as a series of contrasts between selected portions of the dataset. Differences in the patterns of participation were marked by the amount and sequence of argumentation markers and questions, and structural features of extended inquiry. The analyses also revealed when/if students were able to use the system's scaffolding to supports in concert with forms of talk that signal more sophisticated reasoning about theory and evidence. Because this technique represents data at the level of the classroom, it provides a number of insights about knowledge building as a community activity that would not be apparent by focusing on individual students.

Like the authors in the two previous chapters, Law and her colleagues have used visualizations to represent the chronology of learning activities. By using various methods to generate machine coded data, these authors have resolved some of the labor challenges inherent to capturing the complexity of collaborative learning. This paper can serve as a excellent resource for researchers new to tools for automated coding of discourse and visualization tools. The particular visualization tool demonstrated in this chapter is somewhat more limited in scale than Hmelo-Silver's CORTRA diagrams, but the application of existing automated coding tools signals a way forward for analysts who find themselves overwhelmed with data to code. Several of the tools used were specific to Knowledge Forum data, the Analytic Toolkit and the Knowledge Space Visualizer, but the Visual Intelligent Content Analyzer and the Bobinette tool can be applied to other coded datasets. The authors end with a promise of experimenting with data-mining algorithms to aid the discovery of patterns in knowledge building discourse. If successful, this will add another tool to the growing arsenal of new ways to capture and analyze collaboration.

6.5 Designing to Support (and Capture) Collaboration

The chapter by Stahl presents an ambitious goal: to create a "new science of group cognition" that recognizes that CSCW and CSCL are "fundamentally different from other domains of study". The need for such a science and the basis on which Stahl claims it is necessary is likely to generate lively debate, both within and outside the CSCW and CSCL communities. Rather than joining in the debate, I will restrict my comments here to the methodology presented in this chapter as it is applied to the data from Stahl's Virtual Math Team (VMT) Project.

The Virtual Math Team Project was designed to allow a careful investigation of how learning takes place in groups by restricting the object of inquiry to small groups of students using a text-based online system to talk about math problems. Students working together within the VMT environment do not know each other and are limited to interacting via online chat and shared whiteboards. Stahl positions this as a very intentional design decision that allows the analyst to focus on "the disembodied online realm of group cognition" thus stripping away the need to consider individual and social variables or to worry about coding complex real-time interactions that take place in a larger social context. He further proposes that this restricted interactional environment predetermines the unit of analysis to be at the group level and provides the necessary constraints to ensure the analyst's objectivity. Stahl has used data gathered from this setting to demonstrate various features of collaboration in a number of prior publications.

To conduct analyses of the students' activity in VMT, Stahl has created a "VMT Replayer" that provides a VCR-like interface for replaying sessions. Interpretation of the online data is conducted by groups of researchers/analysts who get together (face-to-face) to discuss interpretations of the interactional work represented in the logged discourse. The analysts' hypotheses about the students' activities are offered and refined through iterations of replayed data, using timing data and examining the relationships between the chat and the whiteboard use to further understand the students' meaning making process. The data analysis continues until the research group reaches consensus or when further analysis provides a more developed characterization of the events. In addition, Stahl makes the session logs available with his published papers so that other researchers can replay session data to make their own judgments about the conclusions offered.

Following the four previous chapters, readers will notice that Stahl's method is absent of coding schemes, statistical analyses, and diagrammatic representations of the data. Rather, the primary form of analysis presented in this chapter relies on discussion between a number of analysts who have license only to frame their interpretation of activity based on what appears in the chat and on the shared whiteboard. This requires considerable restraint, as it is far easier to presume to know what the students are thinking or trying to do than to limit the analysis to the observable evidence. The procedural aspect of Stahl's analysis follows Jordan and Henderson's (1995) method of Interactional Analysis, typically conducted on video data. As proposed in their seminal paper, "The goal of Interaction Analysis, then, is to identify regularities in the ways in which participants utilize the resources of the complex social and material world of actors and objects within which they operate" (Jordan & Henderson, 1995, p. 41). Stahl states that his research group has indeed identified regularities across the data sessions, finding "the same kinds of moves occurring in case after case we analyze." These regularities, he proposes, can lead to a science of group interaction by generating theoretical categories, conceptualizations, structures and principles for understanding group learning.

Like the other researchers represented in this section of this book, Stahl proposes a methodology for understanding how collaborative learning is accomplished. He does so using a process that is no less challenging for the analyst than developing a complex coding scheme, although it does not produce the same kind of data artifacts as those of Hmelo-Silver or Law. Stahl's data is similar to other authors who have focused on online talk conducted in specifically designed learning environments, but his lens is much more focused. It may be too focused to reach his larger goal, however, as Jordan and Henderson (1995) caution, "artifacts and technologies set up a social field within which certain activities become very likely, others possible, and still others very improbable or impossible" (p. 41). By concentrating on students who do not know each other, never meet face-to-face and only communicate online, Stahl may be missing an opportunity to discover the kind of nuanced properties of collaboration that can be revealed by methods presented in the other chapters.

Researchers looking for lessons from this chapter may want to consider using Stahl's approach in combination with other methodologies. For example, one might use Stahl's method to help develop a coding scheme and then using it again after generating interpretations of the data based on that coding to test them against what is clearly observable. Previous work using Interaction Analysis has shown that rigorous standard for interpretation proposed by Stahl can be successfully applied with very rich data, such as video of face-to-face interaction as well as annotated transcripts of talk mediated by technology (telephone and online discussion). Further, this style of analysis can be done with data that is deeply contextualized without necessitating unsupported conjecture about what is happening (see Goldman et al. 2007).

6.6 Where Do We Go from Here?

When asked to write this commentary, the editors requested that I look beyond these chapters to discuss what else we need to focus on to move the methodologies forward. The range of methods displayed by these first five chapters already provides a dizzying array of options to consider for examining collaborative learning, and no doubt the subsequent chapters in this volume will provide more. Rather than propose yet more techniques or specific analytic frames, I recommend an approach taken by a group of researchers who are attempting to build common ground between theoretical approaches and methodologies within CSCL.

This endeavor, lead by Kristine Lund (University of Lyon), Nathan Dwyer (SRI), Carolyn Rose (Carnegie Mellon), Nancy Law (University of Hong Kong), Dan Suthers (University of Hawai'i), and Chris Teplovs (University of Toronto), consists of a series of workshops where researchers share data sets for examination by the workshop group, rotating roles of data presenter, analysis presenter, and discussant. Data presenters describe the dataset, how it was conducted, and the original research question that drove the data collection. Prior to the workshop, analysis presenters have access to a selected dataset and present their own style of analysis of those data. Discussants provide yet another perspective on the analyses and the conclusions that can be drawn from each perspective. The researchers have met four times to iterate through this process: first at the 2008 ICLS meeting in Utrecht, the second at the CSCL 2009 meeting in Rhodes, the third at the Alpine Rendez Vous sponsored by the STELLAR Network of Excellence (see www.stellernet.eu), and a fourth workshop at the 2010 ICLS conference in Chicago. See Lund (this volume) for some of the insights generated by the workshop participants.

I believe that CSCL researchers can benefit from utilizing multiple methods and sharing datasets as Lund and her colleagues have done. The goal, however, is not to create one monolithic methodology that is expected to fit all datasets and theoretical perspectives. Rather I believe that each method can be improved by challenges from researchers with different theoretical and methodological approaches who can question assumptions and illuminate potential errors in perspective. We can benefit from looking across datasets to see where differences in coding (unit of analysis, terminology, representation), context (classrooms, after-school clubs, etc.) and data form (video, audio, text) may obscure similarities in the collaborative process. Commitment to this agenda doesn't require that we give up our own point of view, but only that we be interested in finding how productive discourse between analysts can advance our science. Just as we see that learners can benefit from working together, so can we.

References

Azmitia, M. (1988). Peer interaction and problem solving: when are two heads better than one? *Child Development, 59*(1), 87–96.

Burtis, J. (1998). The Analytic Toolkit. The Ontario Institute for Studies in Education. Knowledge Building Team, The University of Toronto.

Dillenbourg, P. (1999). What do you mean by collaborative learning? In P. Dillenbourg (Ed.), *Collaborative-learning: Cognitive and computational approaches* (pp. 1–19). Oxford: Elsevier.

Dillenbourg, P., Baker, M., Blaye, A., & O'Malley, C. (1996). The evolution of research on collaborative learning. In E. Spada & P. Reiman (Eds.), *Learning in humans and machine: Towards an interdisciplinary learning science* (pp. 189–211). Oxford: Elsevier.

Gee, J. P. (2005). *An introduction to discourse analysis: theory and method*. London: Routledge.

Gee, J., & Green, J. (1998). Discourse analysis, learning and social practice: a methodological study. *Review of Research in Education, 23*, 119–169.

Goldman, R., Pea, R., Barron, B., & Derry, S. (Eds.). (2007). *Video research in the learning sciences*. Mahwah: Erlbaum.

Jordan, G., & Henderson, A. (1995). Interaction analysis: Foundations and practice. *The Journal of the Learning Sciences, 4*(1), 39–103.

Law, N., Lu, J., Leng, J., Yuen. J., & Lai, M. (2008). Understanding knowledge building from multiple perspectives. Paper presented in the Workshop on common framework for CSCL interaction analysis, International Conference for the Learning Sciences 2008, Utrecht, Netherlands.

Law, N., Yuen, J., Huang, R., Li, Y., & Pan, N. (2007). A Learnable Content & Participation Analysis Toolkit for Assessing CSCL Learning Outcomes and Processes. In C. Chinn, G. Erkens, & S. Puntambekar (Eds.), *Mice, Minds & Society: Proceedings of the International Conference on Computer Supported Collaborative Learning 2007* (pp. 408–417). Rutgers University, New Jersey: International Society of the Learning Sciences, Inc.

Linell, P. (1998). *Approaching dialog: Talk, interaction and contexts in dialogical perspectives*. Amsterdam: John Benjamins.

Lui, L. (2008). Trajectories of collaborative scientific conceptual change: Middle school students learning about ecosystems in a CSCL environment. Unpublished Dissertation, New Brunswick.

Luckin, R., Plowman, L., Laurillard, D., Stratfold, M., & Corben, S. (2001). Narrative evolution: Learning from students' talk about species variation. *International Journal of Artificial Intelligence in Education, 12*, 100–123.

Nisbett, R. E., & Ross, L. (1980). *Human inference: Strategies and shortcomings of social judgement*. Englewood Cliffs: Prentice-Hall.

Perret-Clermont, A. N. (1980). *Social interaction and cognitive development in children*. London: Academic.

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

Scardamalia, M., & Bereiter, C. (2003). Knowledge building environments: extending the limits of the possible in education and knowledge work. In A. DiStefano, K. E. Rudestam, & R. Silverman (Eds.), *Encyclopedia of distributed learning* (pp. 269–272). Thousand Oaks: Sage.

Teasley, S. D., Fisher, F., Weinberger, A., Stegmann, K., Dillenbourg, P., Kapur, M., & Chi, M. (2008). Cognitive convergence in collaborative learning. In *Proceedings of the International Conference of the Learning Sciences* (Vol. 3, pp. 360–367). Utrecht, Netherlands, LEA.

Teplovs, C., & Scardamalia, M. (2007). Visualizations for Knowledge Building. Paper presented at IKIT Knowledge Building Summer Institute 2007, Toronto, Canada.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*, 1124–1130.

Weisband, S. (1992). Group discussion and first advocacy effects in computer-mediated and face-to-face decision making groups. *Organizational Behavior and Human Decision Processes, 53*(3), 352–380.

Part II
Understanding Learning within Groups

Chapter 7
Analyzing Collaborative Processes and Learning from Hypertext Through Hierarchical Linear Modelling

Agni Stylianou-Georgiou, Elena Papanastasiou, and Sadhana Puntambekar

Abstract The purpose of this study was to investigate whether supporting sixth grade students to monitor and regulate their group navigation behavior while reading from hypertext would lead to a rich understanding of domain knowledge. Metanavigation support in the form of prompts was provided to groups of students who collaboratively used a hypertext system called CoMPASS to complete a design challenge. Multilevel analysis techniques were used to understand how the provision of metanavigation support to groups interact with group navigation behavior and learner's metacognitive awareness of reading strategies to affect individual learning. The findings of this study revealed that providing metanavigation support to the groups contributed positively in enabling students to gain a rich understanding of domain knowledge. Our findings also indicate that there was a significant negative interaction of students' metacognitive awareness and perceived use of reading strategies and the presence of metanavigation support while interacting with hypertext.

7.1 Introduction

In recent years different methodological approaches have been used to measure and analyze collaborative processes while learning in technology-supported settings. Even though there has been a lot of diversity in the approaches that have been adopted to analyze such complex processes, CSCL research has mostly focused on analyzing group discourse. We agree with Naidu and Jarvela (2006) that there is a need to move beyond focusing only on such analyses and direct attention toward understanding how critical attributes of CSCL contexts interact with group collaboration as well as with individual attributes of collaborative learners. All these factors affect the types of interactions and the learning outcomes in a collaborative

A. Stylianou-Georgiou (✉)
Department of Education, University of Nicosia, Nicosia, Cyprus
e-mail: stylianou.a@unic.ac.cy

S. Puntambekar et al. (eds.), *Analyzing Interactions in CSCL*, Computer-Supported Collaborative Learning Series, DOI 10.1007/978-1-4419-7710-6_7,

technology-supported setting and need to be taken into account while studying the dynamic process of collaborative learning. Analysis of learning at both the individual and the group unit of analysis is necessary (Stahl et al. 2006). The purpose of this study was to investigate whether supporting sixth grade students to monitor and regulate their group navigation behavior while reading from hypertext would lead to a rich understanding of domain knowledge. Metanavigation support in the form of prompts was provided to groups of students who collaboratively used a hypertext system called CoMPASS to complete a design challenge. Multilevel analysis techniques were used to understand how the provision of metanavigation support to groups interact with group navigation behavior and learner's metacognitive awareness of reading strategies to affect individual learning. The findings of this study revealed that providing metanavigation support to the groups contributed positively in enabling students to gain a rich understanding of domain knowledge. Our findings also indicate that there was a significant negative interaction of students' metacognitive awareness and perceived use of reading strategies and the presence of metanavigation support while interacting with hypertext.

7.2 Purpose of the Study

In collaborative research designs data are collected at different levels. Therefore, learning might be affected by variables both at the level of the individual learner and the group. According to De Wever et al. (2007), "individual learners are influenced by the social group and context to which they belong, and the properties of this group are in turn influenced by the individuals who make up that group" (p. 3). One major challenge is how to address the friction between the individual-level versus the group-level as well as cross-level interactions between variables on different levels and their impact on the outcome variables (De Wever et al. 2007). Recent studies on collaborative learning in technology-supported settings have underlined that there is a "multi-faceted methodological problem" in this area of research (Fischer et al. 2004) and there is a need for more accurate research methods (in terms of validity and reliability) to assess the impact of learning and working in CSCL settings (Valcke and Martens 2006).

Rummel and Spada (2004) argued that in order to "crack" the complex processes that take place in collaborative contexts we need to work towards developing a "methodological toolbox" which "could support an informed choice of appropriate methods of analysis" (p. 23). Quantitative methods such as multilevel statistical techniques could be useful tools when studying the relationships of variables with different levels and units of analysis. Such methods enable researchers to model the dependencies in the data and obtain more reliable relationships between variables of interest. Multilevel modeling has the potential to answer many questions of interest in CSCL research especially when analyzing data with a clustered structure (Strijbos and Fischer 2007). In a collaborative research design there are variables describing individuals (micro-level) and variables describing groups (macro-level).

Multilevel analysis techniques allows researchers to study how an outcome variable is influenced by a nested set of factors: the individual student participating in these sessions, and the group the student belongs to. Such techniques are suitable for tackling the hierarchical nesting, the interdependency and the unit of analysis problem and provide more accurate estimates when analyzing data collected in collaborative research settings (De Wever et al. 2007).

In this study we used multilevel analysis techniques to understand how cognitive attributes of collaborative learners might be interacting with group membership to affect learning. We designed and implemented support for navigation (metanavigation support) in the form of prompts to compel groups to think about the processes they employ while interacting with online science texts and help them monitor and regulate these processes. Our goal was to investigate whether supporting groups of sixth grade students to monitor and regulate their navigation behavior would lead to a rich understanding of domain knowledge.

7.3 Research Context: Integrating CoMPASS in the Science Classroom

This study was a part of an implementation of CoMPASS (Puntambekar 2006; Puntambekar and Stylianou, 2005; Puntambekar et al. 2003; Puntambekar et al. 2001) in sixth grade science classes. During this implementation, students collaborated in groups using CoMPASS as a resource to find information and read about the science concepts and principles that were involved in the unit of 'Simple Machines'.

7.3.1 Technological Affordances of CoMPASS

CoMPASS is a science hypertext system that has two tightly integrated modes of representation: a textual representation of the content units and a visual representation in a form of concept maps. CoMPASS maps are dynamically constructed and displayed with a fisheye view based on the strength of the relationships among concepts, illustrating graphically the relationships among key ideas in the text (see Fig. 7.1). The maps show the local sub network of the domain and where the links lead to, enabling readers to see the relationships among the text units (concepts) and make thoughtful decisions of what paths to follow without getting lost or confused. CoMPASS also supports readers to study a science idea in multiple contexts by changing views (top right of screen in Fig. 7.1).

In Fig. 7.1 the reader has chosen to read about work in pulley. Work appears as the focal concept in the map and the text related to work appears in the right part of the screen. The concepts that are most closely related to work appear larger and

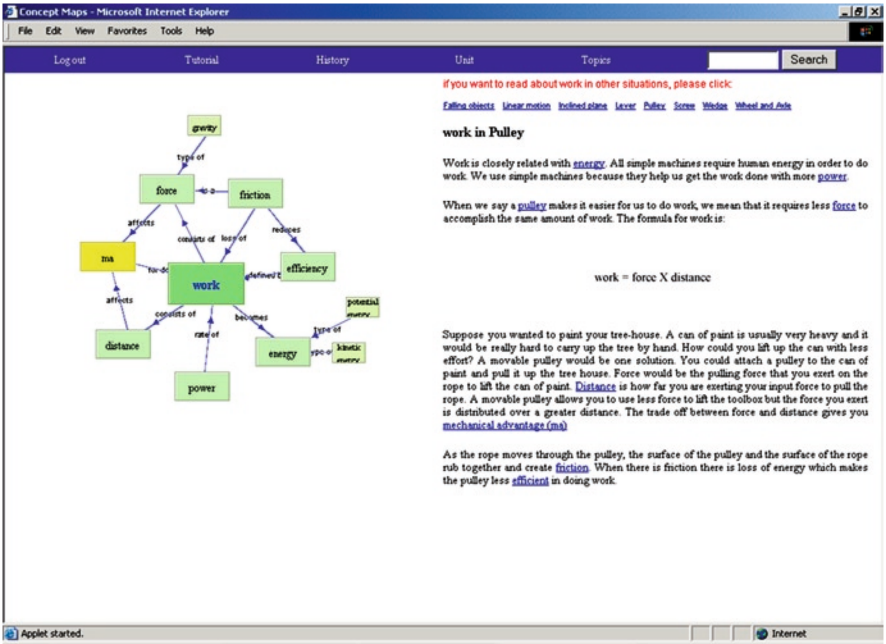

Fig. 7.1 Textual and visual representation of information with 'work' as focus

closer to the focus whereas the concepts that are not as closely related to work appear in the periphery. The maps allow for exploration and support students to take multiple investigation paths based on their learning goals at any particular time.

7.3.2 *Participants*

The participants in this study were 121 sixth graders in four science classes being taught by two different teachers. The school was located in a university town in Connecticut. The students were from different ethnic backgrounds and academic abilities. Each class was randomly assigned to one of two conditions (metanavigation support, no support). Approximately equal numbers of students were assigned to each condition, with variation being due to uneven class sizes.

Students collaborated in groups of three or four while using CoMPASS to solve the "Pulley design challenge". The groups were formed based on teachers' perception of students' academic ability. Teachers decided to form groups of mixed ability levels so that students would benefit from each other during collaboration. The metanavigation support condition included 11 groups of students and the no support condition 15 groups.

7.3.3 Procedures

The study involved four sessions of 45 min that were conducted during the science class period. The *first session* involved an assessment of students' metacognitive awareness and perceived use of reading strategies while reading school-related materials through the MARSI (Mokhtari and Reichard 2002) instrument. This inventory was administered online. The *second session* started with the presentation of the task. The task was a design challenge that required students to build a pulley device that would lift a bottle of water that weighed 600 g off a table using the minimum amount of effort. Students were allowed some time to think about the requirements of the task and write down their initial ideas individually. Then, they were asked to collaborate in groups to plan their quest of finding information to solve the challenge. Groups were asked to read the information that was available for pulleys in the 'Simple Machines' unit in CoMPASS. Groups used CoMPASS for approximately 25 min. During that time group members were reading and discussing the information available in CoMPASS. During the *third session* students were asked to continue their quest of searching information about pulleys in CoMPASS and finalize their pulley system designs. The groups in the metanavigation support condition received metanavigation prompts in a written format to guide their exploration in CoMPASS. Groups were allowed to use CoMPASS for approximately 25 min. It is important to note that while collaborating in groups (session two and three), individual contribution of group members was very important. Group members argued about the navigating choices while using CoMPASS, discussed the information they were reading and how it was relevant to their pulley design ideas. The *fourth session* included an assessment of students' individual science knowledge through a concept map test that was administered in a paper and pencil format.

7.4 Providing Metanavigation Support

Metanavigation support in the form of prompts was provided to the groups in the metanavigation support condition to encourage them to monitor and regulate their navigation strategies in order to gain a rich understanding of science concepts while reading from hypertext. Metanavigation support was based on two indices that were informed by group's navigation path while interacting with the CoMPASS system.

Log file information that captured groups' navigation path enabled us to assess their navigation behavior and decide what metanavigation prompts would be given to each group. Computer log files recorded information about what science concepts the groups explored while using CoMPASS, how much time they spent on each concept and what navigation tools they used to make their navigation choices. Two main indices from group's navigation path informed our decision of what type of metanavigation support each group needed: navigation choices and transitions among text units (see Table 7.1). Specifically we were interested in whether or not the group members had chosen to read about the science concepts that were relevant

Table 7.1 Group navigation based on log file data

Log file information	Type	Description
Concepts visited	Non-goal related Goal related	Do students visit concepts that are relevant to their learning goal?
Transitions	No coherence Coherence	Do students make coherent transitions while reading?

Table 7.2 Conditions for providing metanavigation prompts

	Metanavigation support rules
Navigation choices	If choice of non goal-related concepts ⇒ encourage goal-related navigation
	If goal-related navigation ⇒ encourage integration of science knowledge
Transitions	If transitions are not coherent ⇒ encourage regulation of navigation behavior to make coherent transitions between text units while reading
	If transitions are coherent ⇒ encourage integration of science knowledge

to their learning goal and whether the transitions they made among the text units that were available in the hypertext environment would enable them to gain a rich understanding of the domain. For example, did the group make coherent transitions while reading about science concepts? Coherent transitions were defined as the transitions among closely related concepts in the CoMPASS system.

Considering the binary state of each of these categories, we could have four different cases, described in the "metanavigation support rules" cells of Table 7.2, as well as various combinations. For example, let us consider the situation of a group that chose to read about science concepts that were not important for solving the pulley challenge (i.e., kinetic energy, potential energy, and power) and did not read about goal-related science concepts such as mechanical advantage, distance, and force. The log file information also indicated when the group made incoherent transitions among science concepts while navigating.

The metanavigation prompts that were given to the group were aimed at encouraging students to understand the affordances of the navigational aids in CoMPASS and use them to guide their navigation. The prompts encouraged students to think about their goal and use the concept maps to make thoughtful decisions of what paths to follow. It was pointed out that the concept maps could help students make connections and decide what science concepts were related to what they were reading at any particular time. The metanavigation prompts were distributed to the group members in a written format.

7.5 Data Sources and Measures

Multiple sources of group and individual data were collected over the four sessions. Individual measures included student's individual performance in the Metacognitive Awareness of Reading Strategies Inventory (MARSI) and a concept map test. Group measures included log file information that captured group navigation paths during the use of CoMPASS. These measures were important in order to study how critical attributes of a technology supported setting (provision of metanavigation support to groups while reading from hypertext) interact with group collaboration (group navigation behavior assessed through computer log files) as well as with individual attributes of collaborative students (metacognitive awareness of reading strategies) to affect individual learning outcomes (understanding of domain knowledge assessed through a concept map test).

7.5.1 Individual Measures

The Metacognitive Awareness of Reading Strategies Inventory (MARSI) (Mokhtari and Reichard 2002) was used as a pre assessment instrument to evaluate students' metacognitive awareness and perceived use of reading strategies while reading school-related materials. MARSI consisted of 30 Likert-type items with a 5-point response format (1 = "I never or almost never do this", 2 = "I do this only occasionally", 3 = "I sometimes do this-about 50% of the time", 4 = "I usually do this", 5 = "I always or almost always do this"). An overall total average MARSI score was calculated for each student indicating how often the student uses reading strategies when reading academic materials. This measure was assumed to be an individual characteristic that might have influenced the navigation behavior of the groups.

A paper and pencil concept map test was used as a post assessment tool to evaluate the richness of students' understanding of science concepts. The students were provided with a list of science concepts from which they were asked to create a concept map providing an explanation for each concept, making connections among concepts and stating how they are related. Two aspects of the maps were examined: the explanation provided for the concepts and the explanation provided for the connections among the concepts. Students' concept maps were analyzed using a rubric that was developed in a study conducted by Puntambekar et al. (2003). Students' responses were scored on a scale of 0–3 based on the depth of science understanding that they demonstrated. A score of 0 indicated an incorrect explanation, while a score of 3 indicated a complete and clear explanation for the concept or the connection. A concept ratio was calculated for each student by dividing the score that was given for the explanation of the concepts by the number of concepts included in the concept map. This ratio was a measure of student's understanding of science concepts. A connection ratio was calculated by dividing the score that was given for the explanation of the connections with the number of connections in the map. This ratio was a measure of the depth of understanding of the relationships among science concepts.

7.5.2 Group Measures

Computer log files were used to look more deeply into the navigation paths of groups of learners in an attempt to detect differences in approaches to reading and learning from hypertext when providing metanavigation support. Log files recorded information about what science concepts the groups explored while interacting with the CoMPASS system in a chronological order. Two primary dimensions were used for the analysis of group navigation paths. The first dimension was based on whether groups chose to focus on science concepts that were related with their goal. A goal-relatedness index was calculated by dividing the total number of goal related concepts visited to the total number of concepts visited. The second dimension was based on whether the groups made coherent transitions among the different text fragments. A coherent transition index was calculated by dividing the number of coherent transitions to the total number of transitions among concepts.

7.5.3 Investigations and Data Analyses

The main research question that was addressed in this study was: To what extent can concept maps scores (explanations of concepts and explanations of connections) of students be predicted from the presence of metanavigation support while interacting with science texts, their individual metacognitive awareness of reading strategies and the group navigation behavior?

In order to analyze the data for this study, multilevel analysis techniques were used (Bryk and Raudenbush 1992) with the use of the software HLM 6.01 for windows. Multilevel analysis techniques are helpful for taking into account dependencies that occur in datasets that have hierarchical structures. Accounting for such dependencies is especially important in order to reach reliable estimates of the effectiveness of each independent variable on the outcome variable of interest. For the purpose of the current study, the data were gathered and analyzed on two levels. Level 1 included variables that were gathered on the individual student level; level 2 included variables that were gathered on the group level since the students were nested within groups.

Two-level HLM models were tested on two outcome variables. The first outcome variable was the concept ratio (CONCR), a measure of student's understanding of science concepts. The second outcome variable was the connection ratio (CONNECTR), a measure of the depth of understanding of the relationships among science concepts. For each outcome variable, the HLM analyses were performed in three stages. At the first stage, a null model was tested in which no independent variables were included in the analysis. The results produced by this model were comparable to a random effects ANOVA which measured the variance within and between groups. At the second stage, the student-level independent variables were added to the model, while at the third stage the group-level independent variables were added. The independent variables were added to the model based on theory. However, cross-level interactions that were not significant were deleted from the final models.

The level 1 data included student level characteristics, which were those of the student's metacognitive awareness and perceived use of reading strategies while reading school-related materials (MARSI). The level 2 data included group level characteristics which were those of the condition that the students were in (whether they received metacognitive support or not), as well as the two navigation dimensions that were used for the analysis of group navigation paths. The first dimension was the goal-relatedness index (GOALNAV), a measure of whether groups chose to focus on science concepts that were related with their goal. The second dimension was the coherent transition index (TRANSNAV), a measure of whether the groups made coherent transitions among the different text fragments while interacting with CoMPASS.

Table 7.3 includes a more detailed description of the variables used in the analysis. More specifically some descriptive statistics, such as the means, standard deviations as well as the minimum and maximum values of each variable are presented. As shown in Table 7.3 the average concept ratio score was higher than the average connection ratio score. It seems that students did not provide many complete and clear explanations for the connections among concepts in their concept map (mean = 0.8). The table also shows that the average score of the goal-related navigation index was higher than the average score of the coherent transition index. Groups were better in choosing to read about science concepts that were related with their goal than making coherent transitions among the different text fragments. As far as students' metacognitive awareness of reading strategies is concerned, it seems that on average students reported that they usually apply reading strategies when reading academic or school related material.

Table 7.3 Description of variables used in the models

Name	Description	Level	Type	Minimum	Maximum	Mean	SD
CONCR	Concept Ratio in Concept Map	1	Outcome	0.00	2.75	1.32	.63
CONNECTR	Connection Ratio in Concept Map	1	Outcome	0.00	1.60	0.80	.34
MARSI	Metacognitive Awareness of Reading Strategies Score	1	Predictor	1.30	4.70	3.11	.71
CONDITION	Indicator of whether the groups received metacognitive support or not	2	Predictor				
GOALNAV	Goal-related Navigation Index	2	Predictor	0.00	1.00	0.66	.31
TRANSNAV	Coherent Transition Index	2	Predictor	0.00	1.00	0.57	.28

7.6 Results

7.6.1 Predicting Connection Ratio in the Concept Map Test

The first analysis that was performed wanted to examine the depth of understanding of the relationship among science concepts. This depth of understanding, also called the connection ratio (CONNECTR) was the first dependent variable that was examined with HLM. Equations 7.1–7.3 represent the final model for this sample. Through these models we attempted to explain the differences that students hold in their depth of understanding of relationships More specifically, Eq. 7.1 represents the effects of each student's MARSI score on the CONNECTR variable. This equation examined whether each student's metacognitive awareness of reading strategies had an effect on their depth of understanding of relationships. The parameter β_0 represents the intercept of the dependent variable CONNECTR, which reflects the value of the dependent variable when the MARSI score is equal to zero. The parameter β_0 is further analyzed based on the student's group level variables, as presented in Eq. 7.2. The parameter β_1 represents the amount of change by which the dependent variable CONNECTR increases, for each unit of change of the independent variable MARSI. The parameter β_1 is further analyzed based on the student's group level variables, as presented in Eq. 7.3.

Equation 7.2 represents the group level main effects of CONDITION, TRANSNAV and GOALNAV. This equation examined whether (a) the condition that the students were in (whether they had received support or not), (b) whether each student's group made coherent transitions among the different text fragments, and (c) whether each student's group focused on concepts that were related to their goals, had an effect on their depth of understanding of relationships. Finally, Eq. 7.3 represents the interaction between the condition that each group was in with each student's MARSI score.

Level-1 Model (Student level)

$$CONNECTR = \beta_0 + \beta_1 * (MARSI) + R \tag{7.1}$$

Level-2 Model (Group level)

$$\beta_0 = \gamma_{00} + \gamma_{01} * (CONDITION) + \gamma_{02} * (TRNSNAV) + \gamma_{03} * (GOALNAV) + Uo \tag{7.2}$$

$$\beta_1 = \gamma_{10} + \gamma_{11} * (CONDITION) \tag{7.3}$$

As shown in Table 7.4, the students who were placed in groups with higher levels of goal navigation, also had higher levels of CONNECTR scores ($\gamma02 = 0.324$, $p = 0.013$). This indicates that the students whose groups chose to focus on concepts that were related to their goals had more depth of understanding of the relationships among the concepts. However, the levels of TRANSNAV that the groups held (whether the groups made coherent transitions) did not appear to have any effects on the student's depth of understanding ($\gamma03 = 0.019$, $p = 0.891$). The results of this

Table 7.4 Coefficients of the Connection Ratio Model

Effect	Symbol	Coefficient	Standard error	T-ratio	Approximate df	*p*-value
OVERALL INTERCEPT	β0	0.411	0.160	2.560	22	0.018
CONDITION	β1	0.692	0.242	2.866	22	0.009
TRNSNAV	γ01	0.019	0.136	0.139	22	0.891
GOALNAV	γ02	0.324	0.119	2.713	22	0.013
MARSI	γ03	0.039	0.041	0.948	81	0.346
CONDITION*MARSI	γ11	−0.169	0.068	−2.503	81	0.015

analysis have also shown a significant interaction between the condition that the students were in (whether they had received support or not), with the student's metacognitive awareness (MARSI) ($\gamma 11 = -0.169$, $p = 0.015$). The negative sign of the gamma weight indicates that the students who had received support, but who had lower levels of metacognitive awareness, also had lower levels of depth of understanding. Based on the same relationship, the students who had not received support, but who had high levels of metacognitive awareness also had lower levels of depth of understanding.

In order to determine the percentage of variance explained by the models, it was important to estimate the baseline variance that was accounted for in the null model, when no independent variables are added. Based on the unconditional model, the percentage of variance between groups was 11.09%. As a next step, the level 1 predictor (MARSI) was included in the model. Although this variable did not help explain any of the level 1 variance, it was kept in the model in order to test for its interaction with the condition. However, the addition of the MARSI variable did help explain 15.9% of the variance at level 2. Finally, when the final complete model was run, it was able to explain 3.3% of the variance in level 1, and 99.73% of the variance in level 2.

7.6.2 Predicting Concept Ratio in the Concept Map Test

The procedures that were mentioned above were also performed with the concept ratio (CONCR) as the dependent variable, which measured the student's understanding of science concepts. As a first step, the same complete model that was used above was tested with CONCR as the outcome variable. However, the variables TRANSNAV and GOALNAV were not statistically significant, and were therefore deleted from the model. Equations 7.4–7.6 describe the final model that was used for this dependent variable.

Level-1 Model (Student level)

$$Y = \beta 0 + \beta 1 * (MARSI) + R \tag{7.4}$$

Level-2 Model (Group level)

$$B0 = \gamma 00 + \gamma 01 * (CONDITION) + Uo \tag{7.5}$$

$$B1 = \gamma 10 \tag{7.6}$$

Equation 7.4 represents the level 1 effects of each student's MARSI score on the CONCR variable. More specifically, this equation examined whether each student's metacognitive awareness of reading strategies had an effect on their depth of understanding of science concepts. The parameter β_0 represents the intercept of the dependent variable CONCR, which reflects the value of the dependent variable when the MARSI score is equal to zero. The parameter β_0 is further analyzed based on the student's group level variables, as presented in Eq. 7.5. The parameter β_1 represents the amount of change by which the dependent variable CONCR increases, for each unit of change of the independent variable MARSI. The parameter β_1 is further analyzed based on the student's group level variables, as presented in Eq. 7.6.

Equation 7.5 represents the group level main effects of condition, which demonstrated whether the condition that the students were in (whether they had received support or not) had an effect on their understanding of science concepts. Finally, Eq. 7.6 demonstrates that the effect of the student's metacognitive awareness on the their understanding of science concepts is fixed, meaning that the relationship between metacognitive awareness and the student's understanding of science concepts is the same across all groups.

Table 7.5 describes the effect that each variable had on the dependent variable of interest (CONCR). The independent variable of MARSI was not significant in explaining the student's CONCR scores ($\gamma 10 = -0.100$, $p = 0.153$). This indicates that the metacognitive awareness of the students did not have any statistically significant effect on their understanding of science concepts. However, the condition was significant ($\beta 1 = 0.359$, $p = 0.011$), indicating that the students whose groups had received support, had higher levels of understanding.

In order to determine the percentage of variance explained by this second model, the baseline variance was estimated from the null model, where no independent variables were added. Based on the unconditional model, the percentage of variance between groups was only 7.93%. As a next step, the level 1 predictor (MARSI) was included in the model, which did not help explain any of the variance in any of the two levels. Finally, when the final complete model was run, it was able to explain 0.03% of the variance in level 1, and 96.02% of the variance in level 2.

Table 7.5 Coefficients of the Concept Ratio Model

Effect	Symbol	Coefficient	Standard error	T-ratio	Approximate df	*p*-value
OVERALL INTERCEPT	$\beta 0$	0.842	0.226	3.729	24	0.001
CONDITION	$\beta 1$	0.359	0.129	2.784	24	0.011
MARSI	$\gamma 10$	0.100	0.069	1.442	84	0.153

7.7 Conclusions

In this study we used multilevel analysis techniques to understand how critical attributes of a technology supported setting (provision of metanavigation support to groups while reading from hypertext) interact with group collaboration (group navigation behavior) as well as with individual attributes of collaborative students (metacognitive awareness of reading strategies) to affect individual learning outcomes (understanding of domain knowledge assessed through a concept map test). An overall result that can be concluded from this study is that providing metanavigation support to the groups seems to have contributed positively in enabling students to gain a rich understanding of domain knowledge and have higher scores in the concept map assessment task. The predictive models that were generated using multilevel analysis techniques for both outcome measures in the concept map assessment task, suggest that the variability in concept maps scores (explanations of concepts and explanations of connections) at the group level was accounted for by the presence of metanavigation support. Although the group level variance was very small, for both outcome measures in the concept map test we were able to explain almost all of the group variance.

The variability in the scores regarding the explanations of the connections that each student provided in his/her concept map was accounted by the presence of metanavigation support, the goal related navigation index and by an interaction of his/her MARSI score with the presence of metanavigation support. The presence of metanavigation support and the goal related navigation index had positive significant main effects on the variability of the explanations of connections among concepts in students' concept maps. Students who collaborated in groups that were given metanavigation support and chose to read about concepts relevant to their learning goal gained a deep understanding of the relationships among science concepts. Our findings also indicate that there was a significant negative interaction of students' metacognitive awareness and perceived use of reading strategies while reading from traditional texts and the presence of metanavigation support while interacting with hypertext. If a student had a low MARSI score (reported that he/she is not using frequently reading strategies while reading from traditional texts) the metanavigation support seems not to have helped them gain a rich understanding of the domain. Also students who had a high MARSI score but were not provided with metanavigation support did not gain a rich understanding of the domain. Providing metanavigation support to groups whose members reported more frequent use of reading strategies might have stimulated collaborative interactions which led to deeper understanding of the relationships among science concepts. This finding supports the claim made by Teasley et al. (2009) that collaborative processes do not necessarily lead to better individual outcomes nor guarantee that all groups will demonstrate the same learning outcomes. Therefore, it is important to study the interrelationships among individual attributes of collaborative learners, group level factors as well as critical attributes of collaborative research contexts in order to assess the learning outcomes in CSCL settings. Another finding of the study was that the models that were created using the multilevel analysis techniques were not effective in explaining the variance on the student level. The MARSI score was not a significant predictor of students' performance in the concept map test. Other variables need to be used to predict the variance at the individual level. Reading comprehension and prior domain knowledge were found to be significant predictors of students' understanding of domain knowledge when we used regression analyses (Stylianou and Puntambekar 2004). In this study we chose to add the MARSI variable at the student level because we were more interested to explain the group variance and specifically what is the nature of this variable (metacognitive awareness of reading strategies applied while reading from traditional texts) with the group measures (group navigation behavior and provision of metanavigation support to the groups).

Overall, applying Hierarchical Linear Modelling enabled us to model the dependencies in the data (in our case students within groups) and obtain reliable relationships among the variables of interest. Moreover, an important pedagogical implication was uncovered using this novel methodological tool, since not all groups who received metanavigation support demonstrated the same learning outcomes. We argue that multilevel analysis techniques can help us unravel some aspects of the complex collaborative processes that take place in a technology-supported setting. It is important, though, to study collaborative processes from multiple perspectives (Hmelo-Silver 2003; Rummel and Spada 2004) and apply different methodological approaches to understand the complexity of interactions and learning in such dynamic contexts. We suggest that new techniques such as multilevel modeling could be combined with known ones that have been widely used in CSCL research to study the relationships of variables with different levels and units of analysis.

Our future research plans are to "crack" the collaborative interactions of groups by examining audio data of peer interactions during navigation. We plan to focus on groups whose members which had high MARSI scores but not given support and investigate the negative interaction in the connection ratio predictive model. We will attempt to understand the richness of information contained in a collaborative interaction and identify what aspects characterize good collaboration which might lead to in-depth understanding of domain knowledge. Such analyses can contribute to our understanding of the reading comprehension processes employed while interacting with hypertext. Identifying how readers navigate digital texts and what kind of support they need while processing nonlinear information will be an important contribution in the hypertext as well as the literacy research fields.

Acknowledgments This research is supported by National Science Foundation's early career grant (CAREER #9985158) to Dr. Sadhana Puntambekar. We thank the students and teachers who participated in the study.

References

Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models*. Newbury Park: Sage.

De Wever, B., Van Keer, H., Schellens, T., & Valcke, M. (2007). Applying multilevel modeling to content analysis data: Methodological issues in the study of role assignment in asychronous discussion groups. *Learning and Instruction, 17*(4), 436–447.

Fischer, F., Weinberger, A., & Mandl, H. (2004). Cracking the nut - but which nutcracker to use? Diversity to approaches to analyzing 23 collaborative processes in technology-supported settings. In Y. B. Kafai, W. A. Sandoval, N. Enyedy, A. S. Nixon, & F. Herrera (Eds.), *Proceedings of the Sixth International Conference of the Learning Sciences – Embracing diversity in the learning sciences* (pp. 23–26). Mahwah: Lawrence Erlbaum.

Hmelo-Silver, C. E. (2003). Analyzing collaborative knowledge construction: Multiple methods for integrated understanding. *Computers & Education, 41*, 397–420.

Mokhtari, K., & Reichard, C. A. (2002). Assessing students' metacognitive awareness of reading strategies. *Journal of Educational Psychology, 94*(2), 249–259.

Naidu, S., & Jarvela, S. (2006). Analyzing CMC content for what? *Computers & Education, 46*(1), 96.

Puntambekar, S. (2006). Learning from digital texts in inquiry-based science classes: Lessons learned in one program. In *Proceedings of the seventh international conference of the learning sciences* (pp. 564–570). Indiana, Bloomington.

Puntambekar, S., & Stylianou, A. (2005). Designing navigation support in hypertext systems based on navigation patterns. *Instructional Science, 33*, 451–481.

Puntambekar, S., Stylianou, A., & Hübscher, R. (2003). Improving navigation and learning in hypertext environments with navigable concept maps. *Human-Computer Interaction, 18*(4), 395–426.

Puntambekar, S., Stylianou, A., & Jin, Q. (2001). Visualization and external representations in educational hypertext systems. In J. D. Moore, C. L. Redfield, & W. L. Johnson (Eds.), *Artificial intelligence in education, AI-ED in the wired and wireless world* (pp. 13–22). Netherlands: IOS Press.

Rummel, N. & Spada, H. (2004). Cracking the nut – but which nutcracker to use? Diversity in approaches to analyzing collaborative processes in technology-supported settings. In *Proceedings of the Sixth International Conference of the Learning Sciences* (pp. 23–26). UCLA, Santa Monica.

Stahl, G., Koschmann, T., & Suthers, D. (2006). CSCL: An historical perspective. In R. K. Sawyer (Ed.), *Cambridge handbook of the learning sciences* (pp. 409–426). Cambridge: Cambridge University Press.

Strijbos, J. W., & Fischer, F. (2007). Methodological challenges for collaborative learning research. *Learning and Instruction, 17*(4), 389–393.

Stylianou, A. & Puntambekar, S. (2004). Understanding the role of metacognition while reading from nonlinear resources. Paper presented at the Sixth International Conference of the Learning Sciences (ICLS), Santa Barbara, CA.

Teasley, S. D., Fischer, F., Dillenbourg, P., Kapur, M., & Chi, M. (2009). Cognitive convergence in collaborative learning. Symposium ICLS 2009.

Valcke, M., & Martens, R. (2006). The problem arena of researching computer supported collaborative learning. *Computers & Education, 46*(1), 1–5.

Chapter 8
Analyzing Collaborative Interactions with Data Mining Methods for the Benefit of Learning

Peter Reimann, Kalina Yacef, and Judy Kay

Abstract In this paper, we attempt to relate types of change processes that are prevalent in groups to types of models that might be employed to represent these processes. Following McGrath's analysis of the nature of change processes in groups and teams, we distinguish between development, adaptation, group activity, and learning. We argue that for the case where groups act as activity systems (i.e., attempt to achieve common goals in a co-ordinated manner involving planning and division of labour), the notion of a group process needs to take into account multiple types of causality and requires a holistic formal representation. Minimally, a process needs to be conceived on the level of patterns of sequences, but in many cases discrete event model formalisms might be more appropriate. We then survey various methods for process analysis with the goal to find formalization types that are suitable to model change processes that occur in activity systems. Two types of event-based process analysis are discussed in more depth: the first one works with the view of a process as a sequence pattern, and the second one sees a process as an even more holistic and designed structure: a discrete event model. For both cases, we provide examples for event-based computational methods that proved useful in analyzing typical CSCL log files, such as those resulting from asynchronous interactions (we focus on wikis), the those resulting from synchronous interactions (we focus on chats).

8.1 Introduction

The general goal of this chapter is to introduce concepts and methods for applying computational methods to identify salient interaction processes from log files (largely) automatically. While data mining methods are increasingly used to

P. Reimann (✉)
The Faculty of Education and Social Work, Education Building A35.252,
University of Sydney NSW 2006
e-mail: peter.reimann@sydney.edu.au

S. Puntambekar et al. (eds.), *Analyzing Interactions in CSCL*, Computer-Supported Collaborative Learning Series, DOI 10.1007/978-1-4419-7710-6_8,

discover regularities and trends in human performance data (in particular for understanding consumer behaviour, e.g., Giudici and Passerone 2002) applying these methods meaningfully proofs challenging, for at least two reasons: Firstly, the GIGO (Garbage In, Garbage Out) principle holds; when the quality of the data is bad, or when the data are represented at the wrong level of granularity (with respect to the mining purpose), the outcomes of data mining will be of limited value, no matter how smart the algorithm. Secondly, in order to interpret the output of data mining methods one has to have a certain level of understanding of the underlying algorithms; the output does not speak for itself. This is particularly true for the case of process mining, the focus of this chapter.

In the next pages, we make two contributions. The first is conceptual, reflecting on the relation between the kinds of change processes that are prevalent in groups and the types of models that might be employed to represent these processes. We argue that for the case where groups act as activity systems (i.e., attempt to achieve common goals in a co-ordinated manner involving planning and division of labour), the notion of a group process needs to take into account multiple types of causality and requires a holistic rather than atomistic formal representation (assuming formalisation is intended). Minimally, a process needs to be conceived on the level of patterns of sequences, but in many cases discrete event model formalisms might be even more appropriate. The second contribution is more practical, providing examples for event-based computational methods that proved useful in analyzing typical CSCL log files, such as those resulting from asynchronous interactions (we focus on wikis), the those resulting from synchronous interactions (we focus on chats).

The chapter is structured as follows: We begin with an analysis of the nature of change processes in groups and teams, distinguishing kinds of changes that take the form of development, adaptation, group activity, and learning. We then survey various methods for process analysis with the goal to find formalization types that are suitable to model change processes that occur in activity systems. In the second half of the chapter, we discuss in more depth and illustrate two types of event-based process analysis: the first one works with the view of a process as a sequence pattern, and the second one sees a process as an even more holistic and designed structure: a discrete event model.

8.2 Temporality in Groups: Development and Change

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 group's response to (actual or anticipated) changes in the embedding context" (p. 6); (c) learning processes, which are based on a group's experience and reflection thereof; and (d) the group's operational processes, actions and activities, which are hierarchically and sequentially connected. We can speak, with McGrath and Tschan (and with

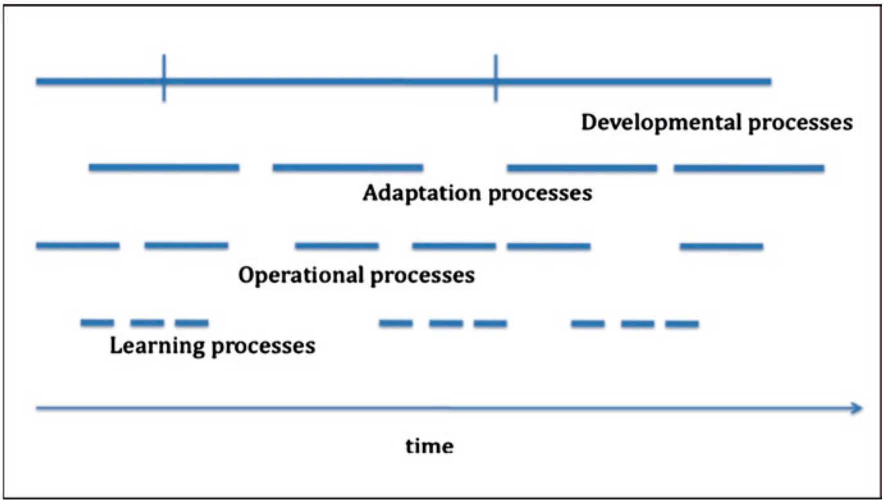

Fig. 8.1 Types of processes effective in groups on different time scales

Aristotle 1941), 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 types of forces are intrinsically temporal, and can operate simultaneously, as illustrated with Fig. 8.1. 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 we 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 view of process introduced below allows us to relax this assumption.

8.2.1 Phases of Group Development

Groups are often portrait as revealing a temporal order by going through a number of *stages or phases* (e.g. Tuckman 1965). It is important from the outset to distinguish clearly between processes that pertain to a group as an *intact system or entity* from *operational processes* that the group performs. Developmental stages or phases a group *goes through* as an entity are of a different kind than the phases a group *brings about* when working on a multi-step task. Developmental processes can be described to a large extent independently from the specific task a group performs, whereas operational processes are much more task-specific. For instance, decision phase models as described by e.g. Poole and Doelger (1986), are not developmental models in the sense intended here.

The prototypical stage theory is (Tuckman's 1965; Tuckman and Jensen 1977) suggestion of *forming, storming, norming, performing*, and *adjourning*. The original theory proposed that all 'normal' groups go through all of these phases over time, and that they do so in this specific order. Another often referenced stage model is Wheelan's (1994), which distinguishes the five stages of Dependency and Inclusion; Counterdependency and Fight; Trust; Work; and Termination. As is the case with most successive-stage models in psychology, strong assumptions about the necessity of each of the phases and about the exact order become disputed, and eventually refuted by empirical evidence (e.g. Poole and Doelger 1986).

Given this fate of numerous stage-based group development theories, McGrath and colleagues (Arrow et al. 2000) have more recently employed concepts from complexity theory, suggesting that what imposes as long-scale group development are not processes that are independent from the micro-interactions in a group, but are global variables that reflect the interactions of local variables (i.e., variables that capture attributes of micro-interactions between group members). This leads to the view that group development on the global level will very much depend on what is going on in a team at the task level, and that hence a search for general development theories is futile, other than in terms of abstract modes such as *formation*, *operations*, and *metamorphosis*. See also Kapur et al. (2007) for an argument for a complexity theory view of large-scale social collectives.

In general, the stratified nature of groups (see Fig. 8.1) makes the detection of invariant sequences at any specific level extremely difficult, because any observable event will be co-determined in multiple ways. For instance, a single contribution to a chat or a wiki reflects at the same time aspects of the development of the group, parts of an adaptation process, a step in an operational process, and perhaps an occurrence of learning going on.

8.2.2 Adaptation to Change

While group development is driven by a group's internal forces, adaptation is driven by events in the group's embedding contexts: "Adaptation refers to changes both in the group as a system and in parts of its embedding contexts, that arise in response to various actual and anticipated actions and events in the embedding systems that will or may affect the group" (McGrath and Tschan 2004, p. 110).

Adaptation processes are well understood in complex systems theory (Sterman 2000) that suggests to see groups as operating in a *fitness landscape* (see also Arrow et al. 2000 from a group research perspective). The external environment of a group is seen as a multidimensional space, with some "locations" in this space offering better payoffs for the group and its members, or having lower costs (McGrath and Tschan 2004, p. 110). Any event in the group's environment, or in the group, can alter the group's location in this fitness landscape, and can change the extent to which the current location is optimal.

From a process analysis point of view, adaptation processes in groups give rise to various challenges. For one, any system in the complex systems theory sense will display diverse time delays between change events and adaption processes, and the relation between quantitative aspects of the change events and of the adaption outcomes will be non-proportional (Sterman 2000). If the system involves humans, such as a group or an organization, the system will be able to *time-shift* responses to change because such systems can anticipate context changes and 'react' to them *before* the actual event occurs, or long after (McGrath and Tschan 2004). McGrath and Beehr (1990) distinguish accordingly between five temporal zones for coping responses: (1) preventive – long before the change event; (2) anticipatory – just before the event; (3) dynamic – simultaneous with the event; (4) reactive – immediately after the event; and (5) residual – long after the event is over. Such characteristics of the relation between cause and effect are hard to align with theories, study designs, and analysis methods that build on a push-type model of causation (Abbott 1988; Monge 1990).

8.2.3 Groups as Activity Systems

To make matters worse, the number of temporal 'issues' grows significantly when we see groups not only as complex systems, but also as *activity* systems—as entities that carry out their projects (Engeström 1999; McGrath and Tschan 2004). Whenever actions are to be carried out collectively, the system undertaking this will have to master several inherent problems, each of which has a key temporal aspect: (1) scarcity of resources, including temporal resources; (2) ambiguity or uncertainty about future actions and events; (3) potential of multiple conflicting interests, including temporally conflicting interests. From these problems inherent to all collective action result a number of requirements to solve them, which McGrath and Tschan (2004, p. 122) summarize as follows:

"The first of these, the inherent scarcity of time and other resources, brings the need for setting priorities. This in turn brings the need for strategic planning, including the selection and acceptance of collective goals or purposes and the allocation of resources to their attainment. The second problem, the inherent uncertainty about future events and actions, brings the need for predictability, and thus for operational planning, including the scheduling of the *who, what, how,* and *when* of various activities required for attainment of those goals. The third problem, the potential for conflicting demands, brings the need for *coordination,* hence for the synchronization both of multiple actions by a given individual and of actions by multiple members of the collectivity."

They further note that although allocation, scheduling, and synchronization are all temporal matters, the three work with qualitatively different *time-reckoning systems*. Allocations are typically expressed in terms of amounts of "staff time", and hence affected by staff members' capacities. Scheduling, on the other hand, is usually done in terms of calendar and clock time, while synchronization and coordination of action are reckoned in terms of an internal or collectivity-defined time.

For instance, members and groups might differ in what they consider to be an appropriate reaction time to a posting on a discussion board.

The complexities of collective action make it less than highly probable that one will find general characteristic sequences of task performance phases, and even less probable that specific sequences are in a systematic manner related to task accomplishment. And indeed, although numerous models have been suggested for specific phases of group problem-solving (e.g., Bales and Strodtbeck 1951 three phase sequence of orientation-evaluation-control) and decision making (e.g. Poole and Doelger 1986), none of them (in their strong form) survived empirical testing. One reaction to this was to not build assumptions of fixed sequencing into theories, but to develop models that hold that groups do show certain phases contingent on task requirements, but not necessarily in a fixed order (e.g. Gouran and Hirokawa 1996).

From a complex systems perspective, the general lack of evidence for specific fixed sequences does not imply that there aren't regular patterns to be found. However, due to the complex multi-layered nature of collective action, they will most likely not occur in the form of specific sequences on the lowest level of the hierarchy (group actions), but will emerge as patterns *across* these levels.

With respect to the question if sequence does matter for improving group performance in problem-solving and decision-making tasks, a consensus has emerged amongst team researchers that any fixed structuring of the task process is likely to lead to better performance compared to no intervention at all, but that no specific reasonable ordering yields additional performance gains (McGrath and Tschan 2004, p. 137). In other words, any reasonable structure (script) provided to groups for coordinating their decision making is better than no structure, but it proofed impossible to demonstrate the advantage of any specific structure. This has important implications for normative recommendations regarding what should be taught to groups to 'optimize' their problem solving and decision making.

8.2.4 Learning from Experience

Learning builds on a group's own history and its own experience. Of course, a group's learning is connected to development, adaptation, and operational activities because these processes are part of the group's history and constitute its experiences. We can sharpen the contrast to development and adaptation by conceiving of group learning as involving an element of *deliberate* processing and reflection. For example, we would speak of group learning--and not simply adapting to changes in the environment--when students in a problem-based learning team engaged in documenting and critiquing their collective reactions to feedback from a tutor.

This element of reflection helps us also to understand a group's *history* as being not simply the sequence of quasi-objective events from its instigation up to 'now', but as consisting of a shared interpretation of what those events were, and what the implications for the group have been up to 'now'. In other words, a group's history is a *resource* for the group—created by the group and at the same time affecting its future decisions. How a group creates its history is not only dependent on the (on-going) interpretation of past events, but also on the *anticipated* future, and the anticipated *group life span*. Hence, it matters if a group is set up (and/or sees itself) as a *task force*, a *crew*, or a *team*. For a task force, the life span consists of the duration of the project that is the main reason for its creation. A task force will always work in the context of "time remaining" to accomplish its mission, with the "half-life" point often leading to dramatic shifts in procedures if not structures (Gersick 1988). For a crew, while the procedures will be mostly stable over time, the frequent changes in crew personnel (e.g. in an airplane, an operation team) leads to seeing crew history as existing only for the duration of the current "shift". Team members, however, can reasonably anticipate on-going collaboration with a more or less stable group of people, and will hence develop a comparatively richly articulated and group-specific interpretation of their history.

We will not delve into group learning deeply because this is, after all, one of the two defining elements of CSCL research, and readers of this book will be well informed. It may hence suffice to mention that according to McGrath and Argote (2001) group learning corresponds to changes in one or more of six *group memory systems*, namely the six sub-networks that make up a group's overall coordination network: member-member relations (social network), task-task relations (task decomposition), tool-tool relations (e.g., web services), member-task relations (division of labor), member-tool relations (roles), and task-tool relations (jobs). Learning can be triggered and supported by processes fully internal to a team, but groups can also learn from other groups, that we can conceive of a kind of transfer of knowledge on the group level (Argote and Ingram 2000). Such transfer can be brought about, for instance, by exchanges of team members across groups, or re-use of processes or tools across groups.

In summary, group learning is temporal in multiple ways: First, as all learning, it unfolds over time. Second, the state of the group is partly dependent on its own actions at earlier times. Third, learning is cumulative in the sense that current knowledge affects the ease of learning something new. Fourth, part of the learning process involves aspects of anticipated future as well as interpretations of the experienced past (McGrath and Tschan 2004, p. 118).

8.2.5 Conclusions

We hope to have convinced the reader that temporality in groups is not just a "method problem", but a theoretical and substantive one as well: Without embedding notions of process and time in theories of group performance and learning, one most likely misses an essential characteristic of groups. That therefore we need to theorize and analyze process in groups, and do so in a manner that takes into account that time takes on various qualities besides calendar time. Besides this general point, which has been widely accepted in CSCL since its move to focus on process (Dillenbourg et al. 1995), the more specific point we want to make is that when groups act as activity systems, accounting for their performance requires to conceptualize coordination processes at a level more holistic than adjacency between any two events. This is for two reasons: Because different sequences of events can realize the same goal, and/or because groups might use their prior experience with going about their work as a resource.

8.3 Methods for Process Analysis

Synthesizing the vast literature on process analysis (e.g., Sanderson and Fisher 1994) we suggest distinguishing between *atomistic* and *holistic* views of process (see Reimann 2007 for a more comprehensive treatment). The main rationale for this distinction is a view of a process either as being made up of *particulars,* the ordering of which is being governed by an underlying law-like process, or a view of process as a *whole*, for instance a plot-like structure. Along this *granularity* dimension, we can distinguish between (time) series analysis, (event) sequence analysis, and narrative methods.

A second important distinction concerns the unit of analysis, which can be *variables* or *events*. Variables are attributes of fixed entities defined by measurement (e.g., with a scale) or by a coding and counting procedure. The decision to phrase research questions in terms of variables and relations between them is a very decisive one, since 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. Combining these two dimensions of Granularity and Unit of Analysis yields a classification scheme for process analysis methods as depicted in Table 8.1 (for more details see Reimann 2007).

As we focus on data mining methods in this chapter, our focus will be exclusively on event-oriented conceptualizations of process.

Table 8.1 Examples for methods classified according to Granularity and Unit of Analysis

Unit of analysis	Atomistic ←	Granularity of process →	Holistic
	Series	Sequence	Narrative
Variable-oriented	Time series analysis	Quantitative parameters of sequences (e.g. length)	Quantitative parameters of narratives (e.g. word frequencies)
Event-oriented	Stochastic modelling (e.g. Markov models)	Optimal matching methods Graphical methods (Cognitive) simulation models Grammar-based methods Process-modelling and –mining	Interpretative methods; Ethnomethodological approaches Conversation analysis; Event structure modelling

8.3.1 Conceptualizing Process in Terms of Event Sequences

To account for change processes in groups, process analysis methods should be able to deal with at least two other kinds of causes in addition to *efficient* cause, 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 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. A reduction to efficient causes is not sensible 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.

That the explanation of human behavior requires considering multiple types of causality is also supported by scholarship on *agency*. Agency (and associated concepts, such as self-hood, will, motivation, intentionality, choice, freedom) is a pivotal concept in social sciences whenever the relation between social 'forces' and individual choice is at issue. In their famous essay "What is Agency?", Emirbayer and Mische (1998) put *temporality* into the centre of their discussion when defining human agency "…as a temporally embedded process of social engagement, informed by the past (in its habitual aspect), but also oriented towards the future (as a capacity to imagine alternative possibilities) and toward the present (as a capacity to contextualize past habits and future projects within the contingencies of the moment)" (p. 963). The concept of "agentic orientation" introduced by Emirbayer and Mische to account for the phenomenon that humans can choose if they at any moment in time are more directed toward the past, present or future is also constitutive for those psychological motivation theories that emphasize the frame of reference people use to analyze their actions (e.g. Gollwitzer 1986).

Viewing a process in terms of the events that make it up provides the necessary space to account for all three kinds of causality: efficient, formal and final. (As we will not 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). While variable- and event-centred analysis can be combined, conceptually they are quite different and these differences are important to keep in mind (Mohr 1982).

8.3.2 Accounting for Event Sequences with Patterns and Models

The event view of process is compatible with a wide range of models of change, ranging from atomistic to holistic ones, as depicted with the horizontal dimension of Table 8.1. On the atomistic end, we find approaches, such as Markov models, that see processes as a series of events governed by a probabilistic regularity. Like variable-based models, Markov models aggregate the history in the representation of the current state: "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). Histories in this view are a kind of "surface reality" (Abbott 1990) that are generated by deeper, underlying probabilistic processes that finds 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.

On the other end of the granularity dimension, we find narrative models of process and change. For narratives, it is characteristic that they have a plot-like structure, involving actors that are imbued with motives and intentions. While narratives can to some extent be formalized (e.g., Abell 1987) they are typically represented in natural language.

We suggest the notion of a *sequence* to occupy the middle ground between atomistic series and holistic narratives. A sequence is more holistic than a series: the overall form of the sequence matters, not only the relation between adjacent elements; but it is less holistic than a narrative, in particular a sequence does not need to refer to motives and intentions, and does not have to be perceived as a plot or story. For instance, when we speak of a decision making process in a group, we refer to a process that has a beginning and an end, comprises a number of sub-steps (events), and a number of constraints on the order of the sub-steps. However, a sequence does not have to have a plot-like structure, and does not have to convey all the details typical for a narrative. Hence, sequences can be seen as conceptualizations of process more granular than series, and less holistic than narratives. (Note that we are not suggesting a strong distinction here, only a heuristically useful one.)

For the rest of this chapter, we focus on the case of sequence analysis because they have the minimal level of structure, of 'wholeness', to render them serious candidates for capturing change processes in *activity systems*: They allow to account for *activities* and *practices* in groups that are (to some extent) planned and (to some extent) coordinated. At the same time, and different from more holistic structures such as narratives, sequence patterns and models can be identified automatically, with data mining methods.

The question that arises next is how observed sequences can be grouped and classified. One way to do this is to look for *patterns*, for *typical* sequences. One way to find patterns is to use optimal matching algorithms based on a similarity measure for sequences such as the number of changes required to transform one sequence into another (e.g. Abbott and Hrycak 1990) or to cluster observed sequences in other ways (Kaufman and Rousseeuw 1990). Another approach for pattern identification is to rely on graphical representations and use visual cues to group sequences into clusters (e.g., Suthers 2006).

Another way to look at sequences is to see them as *generated* by an abstract process – to treat observed sequences as *instances* of a *model*. This is particularly appropriate when the sequences in the log files can be expected to reflect structured group activities, such as resulting from scripted collaboration (Weinberger and Fischer 2006) or from project-based cooperation (Zumbach and Reimann 2003). As we said before, in such cases, we can think of groups as *activity systems*—as entities that carry out their projects, and of a log file as containing at least in parts records of these structured (planned, coordinated) activities.

To make this distinction between pattern and model clearer and to elaborate on practically relevant implications for data analysis, we provide two examples. The first one demonstrates the use of sequence mining in order to identify patterns of activities from log file data of groups cooperating asynchronously through a wiki and a file repository. The second example introduces the notion of *abstract* process models and how they can be identified computationally, illustrating this method with data from synchronous chat communication. Note that the methods are neither coupled with a specific communication mode (synchronous/asynchronous), nor with any specific collaboration technology such as wikis, chat, etc. We use these data sets to illustrate the range of data that can be subjected (meaningfully) to data mining.

8.4 Sequential Pattern Mining in Asynchronous Interaction Data

The theoretical discussions to this point suggest many promising approaches that might be used to find valuable patterns within the electronic traces of user interaction. This section discusses issues associated with the two key technical approaches for identifying patterns of activities, top down and bottom up, as well as the more abstract issues of mining traces that capture notions of the whole group versus analyses that model individuals and their activity within the group. We then discuss some of the key issues in preparing raw data from interaction into a form that is suitable for mining.

8.4.1 Top-Down, Theory Driven Data Mining on Traces of Group Interaction

Given the complexity of group interaction and the variety of potential goals for mining data derived from group interaction, there appears to be considerable merit in taking a top-down, theory driven approach to the data mining. This means that we begin with theories and we identify elements that seem likely to provide value for understanding how the group is operating and for informing improved group operation.

For example, one meta-analysis of group work (Salas et al. 2005) identified five key elements for successful group work: leadership, mutual performance monitoring, backup behavior, adaptability and team orientation. It also identified three coordinating mechanisms needed: shared mental models, mutual trust and closed-loop communication. Operationalisations of these theoretical constructs in terms of observable behavior can be used for formulating patterns to explore by data mining in a top-down analysis. We (Kay et al. 2006) have used this approach to define temporal patterns expected for aspects such as closed-loop communication (McIntyre and Salas 1995) within and across collaboration media, and these did indeed prove useful for identifying problematic group performance.

The top-down approach to data mining for analyzing collaboration data has various advantages, compared to the bottom-up, data-driven approach: Firstly, relations between pedagogical and psychological theory and research are easily established, and there is a clear expectation of the ways that outcomes from the data mining can be used. Secondly the top-down approach is more likely to provide understandable results that can be interpreted by teachers and learners and then be used to inform learning about ways to improve group operation and to monitor subsequent changes. Thirdly, the *absence* of certain patterns in a group's interaction can be identified in this way, which has high diagnostic value for the quality of communication.

However, since most pedagogical and psychological theories are not developed with the goal to be useful for top-down data mining, it remains challenging to relate theoretical constructs that have been found relevant for group communication and learning to the kind of information that is typically available in log files of CSCL tools. There has been some work where students have used collaboration tools with sentence starters, designed from Speech-Act Theory (Searle et al. 1980). For example, Soller (Jermann et al. 2001) used this approach to achieve stronger analysis of group interaction. Of course, in this case, there is a serious interface burden on the group members: they can no longer write in natural prose as they need to pause before each contribution, think about the type of utterance, and select it, and only then can they go on to write their contribution.

However, a serious problem is that many of the theories of group operation are rather difficult to use to automate collaboration pattern recognition since it is not so clear just how aspects of good or dysfunctional group operation would be evidenced in the data. This leads to the other key approaches that support exploration without a driving theory.

8.4.2 Bottom-up Discovery of Patterns

Computer-based methods also make it relatively easy to take a bottom-up, more exploratory approach to analyzing learner and group interactions. Data mining

Table 8.2 Example of data used in sequential pattern mining

Sequence1	<b,a,d,b>
Sequence2	<a,b,b,d,c>
Sequence3	<a,c,d>
Sequence4	<a,c,c,d>

algorithms, for instance, allow analysis of vast electronic traces, perhaps combined with complementary information (such as exam marks), to discover patterns that are not directly visible to the human eye. As discussed previously, an important aspect of learner data is the temporal aspect of its events. A particularly interesting question for understanding the quality of group interaction is whether there are some sequences of events that are more frequent within successful groups than within dysfunctional ones (or vice-versa), and we would expect that these sequences could be traced back to the higher-level, theory-driven, indicators. Sequential pattern mining (Agrawal and Srikant 1995) is an example of an algorithm that can make use of this temporal aspect and to detect frequent sequences of items (here, events) appearing in a sequence dataset a minimum number of times (called support). For instance in the example of four sequences given in Table 8.2, the sequence <a.d> has a support of 4 as it appears in all four sequences (even though there are other items between the a and the d in all but the first sequence). By contrast, <b,d>has a support of 2 as it appears in Sequences 1 and 2.

When we apply this approach to collaborative data, items such as a, b, c, d would represent an atomic and multi-dimensional event (which typically would have the event's author, the time, the type of intervention, and possibly more information). Our sequences therefore represent a succession of collaborative events. So for example, the first sequence might represent four consecutive contributions on a wiki page, the first one and the last ones being made by the same author (b), or it might represent four consecutive events by the same team member, the first one and the last one being done on the same resource (e.g. a wiki page) and the other two being on other resources.

We have built a variant of the Generalized Sequential Pattern algorithm (GSP) and applied it to our collaborative learner data (Perera et al. 2008). The context was a senior software development project where students worked in teams of 5–7 students and interacted over a collaboration tool comprising of a wiki, a task allocation system and a subversion repository (Trac, see http://trac.edgewall.org/). We searched for sequential patterns within groups, as well as within the behaviour of the individual, and classified the results according to the "quality" of the group, based on marks and on teachers' assessment of the group. Results pointed to the importance of leadership and group interaction, and we were able to identify patterns indicative of good and poor individual practice. An example of interesting patterns found was that groups which were performing well had far more patterns with consecutive events with the task allocation system and the subversion repository (particularly alternating between these), suggesting a high use of the task allocation system and frequent commits to the repository. In contrast, weaker groups displayed the opposite behaviour and in particular had a notable lack of sequences containing subversion events. This was even stronger when we compared leaders' behaviors. This suggests that the work of the group leaders strongly influences the success of the groups. In addition, the patterns showed that these group leaders were much less involved in technical work, suggesting that work in these groups was being delegated properly and the leader was leading rather than simply doing all the work. In contrast, the leaders of the poorer groups either seemed to use the wiki (a less task focused medium) more than the task management system, or they were involved in too much technical work. Looking more closely at sequences of events occurring for a particular task, the better groups had many patterns where the group leader creates and allocates a task to a group member, and this group member formally accepting the task. Various events followed on the wiki or the subversion repository (suggesting that work for that task was being carried out) and then the group member comments and closes the task. A last example, this time looking at the sequences of events occurring on a particular resource, the top group had a large number of patterns where the group leader interacted with the other group members on a specific resource, suggesting a high level of leadership, mutual performance monitoring and close-loop communication. For more comprehensive details about the results found, see Perera et al. (2008).

8.4.3 Synergies from Modeling Groups and Individuals

There are two key approaches for mining logs of group interaction data: at the level of the whole group or the individual. The more obvious approach is the former, where the traces of interaction for the whole group are used. This has some serious limitations because many data mining techniques require large numbers of groups. For example, clustering techniques are intended to identify subgroups in large sets. Yet data mining in CSCL needs to be able to provide useful results for data sets from a single class.

Hence, we took the rather less intuitive approach of modelling individuals' sequences. In hindsight, this is a very promising approach for several reasons beyond the pure pragmatics of data set sizes. Prime among these is that many important changes in behaviour and learning progress occur at the individual level. Put differently, if we can mine the log of each individual's activity within the group and report useful information to the individual, pointing to ways to change their own behaviours to make the group more effective, and being able to track improvements in that individual's behaviour over time, this is really useful for the individual. Another key benefit of mining at the level of the individual is that there are different roles within a group. In particular, the person responsible for leadership needs to fulfill particular roles and data mining that informs them about how to improve their performance as a leader is very valuable for them. Finally, there is the very pragmatic fact that data mining at the level of the individual essentially provides more data points for the data mining. So, for example, our clustering of individuals could be more effective and provide more useful and meaningful results than group mining.

8.4.4 Abstraction from Data Traces to Meaningful Sequences

Challenges arise to prepare the data so that mining can be carried out. The way it is summarized and aggregated has a profound impact on the results and their interpretation. There is a long, complex but critical process of manipulating the data and extracting useful information. Firstly, raw data has very little meaning. The fact that a user clicks on a resource or types in some text does not mean much in itself. But if we know that all other group members also accessed and amended the same resource within the same hour then this could be an indication that the members are sharing information and interacting with each other. Secondly, raw log entries are often too detailed. While a user may think in terms of a whole task, involving work on different resources (wiki and repository entry for instance), these appear as independent resources in the log file.

Hence, prior to performing any kind of bottom-up analysis, it is important to add interpretation and processing to the data so that the subsequent mining has meaning. The evolution of the data can be illustrated as follows (Fig. 8.2):

From the raw data, the first step is to abstract the essential information that is required: for example who did what, when, and where. There are numerous ways to aggregate the data. In our case, we explored the following possibilities, each capturing some aspect and giving different insights on group interaction: (1) an aggregation per resource (to find patterns that occur frequently on a given resource), (2) one per activity session (to find patterns that occur frequently within successive time windows) and (3) one per task (to find patterns that occur frequently in relation to a given task). We designed various *alphabets* to represent our raw data in each of these aggregation methods so that we could encode it effectively, like a language for representing the raw data and abstracting the essential aspects (Perera et al. 2008). Table 8.3 below illustrates the form that data takes along the process.

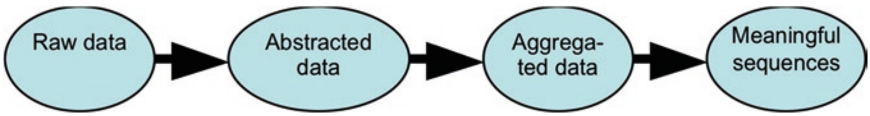

Fig. 8.2 Data interpretation and processing

Table 8.3 From raw data to a meaningful sequence: an illustration

Raw	Abstracted	Aggregated	Meaningful sequence
Very detailed log	User X, Wiki contribution on page Y, Time T	Action sequence on page Y: User X (leader), User Y, User Y, User Z, User X (leader)	Group leader creates a page, 2 other group members contribute and leader contributes again

In addition to deciding what level of data granularity is most appropriately described (with respect to a particular analysis/mining goal), the researcher needs to decide what level of granularity the process model should represent the data. Sequence patterns as described above are one option; another, coarser and more abstract, involves use of discrete event models, to which we now turn (see also Table 8.1).

8.5 Mining for Process Models Based on Chat Data

A discrete event model is a compact, yet comprehensive formal description of all those event sequences that are compatible with the model. Discrete event models are frequently used in theoretical computer science to describe computational processes and in Business IT to describe processes in organisations (e.g. supply chain logistics). They have in our opinion a number of characteristics that make them interesting to add to the repertoire of process analysis methods in CSCL. (Note that we use capitals for Process Model and for Process Modeling in order to distinguish this specific approach from the general notion of process models–which can take many forms, amongst them Process Models.) 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. A defining characteristic of Process Modeling is that log file data are seen as being *generated* by a process (in general, this can be multiple processes, but to keep the explanation concise will speak of one process only) and that this underlying process can be modeled as a discrete event system. More precisely, the log file is interpreted as a sequence of activities that result from (typically) multiple enactments of a process; these enactments form the process *instances*. Since the events in the log file correspond to a more or less small number of activity classes, they can be described with a limited vocabulary.

Figure 8.3 illustrates the situation that is typical for CSCL: In the Collaboration Environment one or more collaboration processes are enacted, supported by a software system such as a chat tool or an argument editor or a more general Learning Management System such as LAMS (www.lamsfoundation.org). The system records the interactions and transactions to some extent in an event log. The event log is used for Process Model mining.

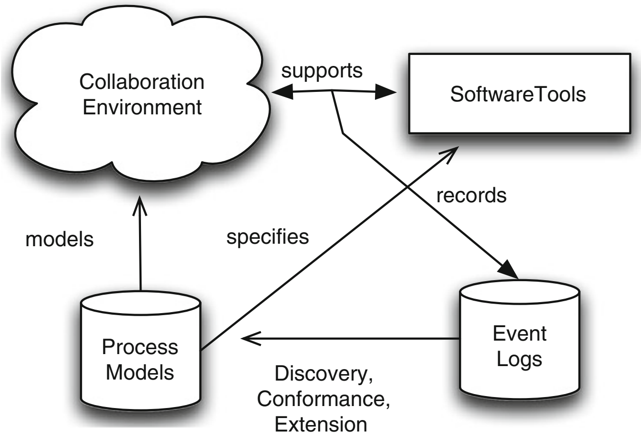

Fig. 8.3 Elements of process mining for CSCL research (Modified after Van der Aalst and Günther 2007)

Process mining can serve a number of purposes, among them:

- Discovery – No a-priori model exists. A model is constructed based on an event log;
- Conformance – An a-priori model exists. Event logs are used to determine the extent to which the enacted collaboration corresponds to the model.
- 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 in the context of collaborative learning. Before we explain how models can be discovered, we need to detail what form they take.

8.5.1 Discrete Event Models as a Model Class for Activity Systems

The class of processes models we 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 events 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,

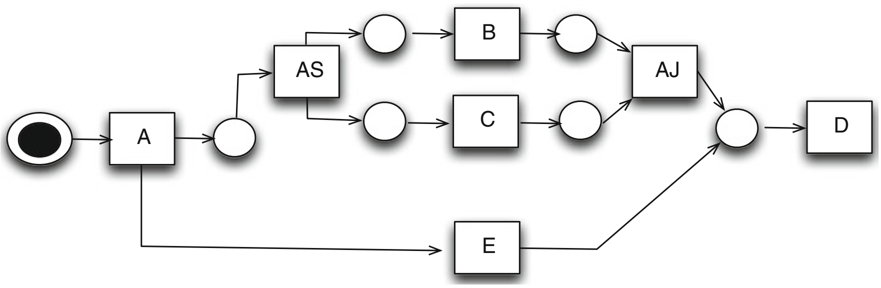

Fig. 8.4 Example for a Petri net description of a process

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. 8.4 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 represents a token, which enables the transition A to be fired. Petri nets are non deterministic 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, an *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. And 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, i.e. serve as a mirroring or feedback device (Muukkonen et al. 2007).

Even a simple Petri net is a basic, but powerful language to represent for instance the logic of a group script, including concurrency of activities. 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. They are used for instance to mine the underlying process from users' interactions with a document management system (Kindler et al. 2006). We explain this contextualized in terms of CSCL research.

8.5.2 Discovering Discrete Event Models

To provide an example for how process mining can be applied to observations from a typical CSCL scenario, we refer to Reimann et al. (2009) and their analysis of decision making events as they occurred in groups of students who worked on a design task using a chat tool for communication. Chat data were first inspected (by a human rater) for all occurrences of a generic decision making process, and the steps involved in each decision instance were then coded by human raters using a slightly modified version of the Decision Function Coding Scheme (DFCS, Poole and Holmes 1995). The DCFS provided the vocabulary for describing the activity classes that can occur in a decision making process (under this theoretical perspective). Looking at one of the groups here, its event log consisted of 23 decision process instances, with a total of 1,115 events occurring in these instances. This group reached their decision in six events in one case, and the longest decision took 234 events. On average, it took 48 events for this group to reach a decision. Table 8.4 shows the DFCS categories and their frequencies in this group.

The transformation of data from a log file into a Process Model representation cannot, at this stage, be fully automatized unless the event data come from highly structured workflow environments. For the kind of data typical for CSCL research, such as chat protocols, 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 Petri nets with a formal semantics are regrettably not suitable, among other reasons because they are overly deterministic. This model class does not easily fit to data that contain noise (i.e., not all events can be seen as belonging to the model) and/or are incomplete (not all model elements of the model are observed at least one time). What is needed for noisy data and incomplete data is a model type that makes less strong assumptions on the relation between events observed and relations in the model. One such model class are *dependency graphs*. In order to mine for dependency graphs, one will have to employ *heuristic* methods.

Table 8.4 Counts of decision event types

	Group A	
Function	N	Frequency (%)
Problem definition	64	5.7
Orientation	512	45.9
Solution criteria	42	3.8
Solution alternatives	130	11.7
Solution elaboration	64	5.7
Solution evaluation (positive)	27	2.4
Solution evaluation (negative)	5	0.4
Solution confirmation	29	2.6
Non-task	146	13.1
Simple agreement	91	8.2
Simple disagreement	5	0.4

8.5.2.1 Heuristic Process Mining

In order to generate a model that summarizes the 23 decision process instances that we were able to identify in the chat data, we (Reimann et al. 2009) used the *HeuristicsMiner algorithm* to discover models from event logs (Weijters et al. 2006). The HeuristicsMiner uses a frequency based metric to express the degree of certainty of a dependency relation between two events *A* and *B* based on an event log *W*, expressed as: $A =>_w B$. With $|a >_w b|$ standing for the *number* of times *a* is followed by *b*, the metric is calculated as:

$$A \Rightarrow_W B = \left(\frac{|a >_W b| - |b >_W a|}{|a >_W b| + |b >_W a| + 1} \right)$$

In words: The number of times a is followed by b is subtracted from the number of times a follows b and this difference is divided by the sum of these two relations, plus 1. This metric takes values between 1.0 and − 1.0, with a value close to 1.0 indicating a high certainty that b follows a, and values close to − 1.0 an almost definite certainty of the reverse (a follows b).

Instead of using a fixed value for $a =>_w b$ as the threshold, the heuristic takes the highest score to decide which relation to put into the dependency graph is appropriate if we request that all observed activities should be connected. The HeuristicsMiner algorithm can deal not only with noisy and incomplete event logs, but also with short loops (e.g. ACCB, ACCCB) and with non-free-choice situations: in some process models the choice between two activities depends on choices made in other parts of the process model.

The result of applying this algorithm (with specific parameters set as described in Reimann et al. 2009) takes the form of a dependency graph depicted in Fig. 8.5. The arcs on the right side of the boxes that point back at their own box indicate loops, meaning that statements of this type often occurred multiple times in a row. The numbers along the arcs show the dependency of the relationship between two events, as explained previously. The second number indicates the number of times this order of events occurred. The numbers in the boxes indicate the frequency of this event. Note that this representation is very similar to the transitional state diagram described in Jeong et al. (this volume).

Space constraints do not permit us to go into the details of this specific model. Instead, let us mention some general points. It is important to note that the Process Model discovered from the (coded) chat transcripts and taking the form of Dependency Graph as displayed in Fig. 8.5 is a model of *all* the 23 process instances occurring in the chat transcripts; it is an aggregation, or generalisation, of these observed instances, but not one using variables. Since the model is discovered using heuristics, the decision if newly observed decision instances (for instance, from a different group of students, or from the same group of students at a different time) is commensurate with the model can not be made in an deterministic manner (as is the case for Petri Nets), but would need statistical methods.

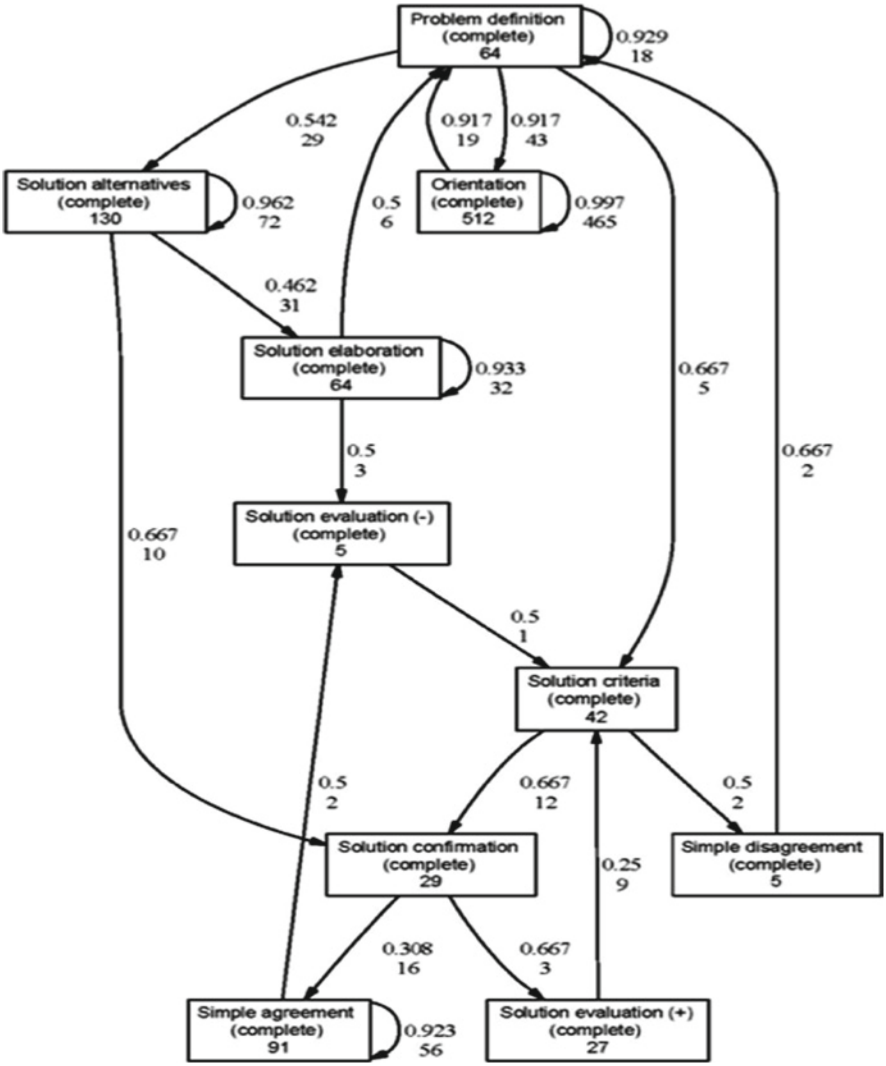

Fig. 8.5 Dependency graph as a visual representation of a group decision making process

8.6 Discussion

We have argued that for groups as activity systems, mining log files with the goal to identify sequence patterns and/or discrete event models in order to capture change processes is a promising approach: not only can patterns and models be discovered automatically in many cases, patterns and in particular discrete event models are also on the right level of granularity for capturing (partially) designed processes that might be used by groups as a resource. However, we have said little so far on how patterns and models relate to learning, as different from task-related activity sequences.

Intentional learning from experience, as described in the introductory section, can be represented with patterns and models in two main forms: (a) A pattern/model can have a direct relation to learning if the activity classes included in the pattern/model are themselves considered to reflect aspects of learning. For instance, events might comprise activities such as Raise Questions, Read Texts, Reflect On Text. A pattern description or discrete event model for such event sequences can be considered to represent orderings of such more or less direct learning activities. (b) Indirect relation to learning: In this case, learning takes the form of *changes* in the pattern/model over time; for instance, a group changes over time how it goes about its decision making or about coordinating its collaborative writing activities. Here we need further to distinguish between change processes that can be described as changes in the ordering of a set of activities, as different from changes that can only be captured by extending or changing the vocabulary for activities. Conceptually, this kind of "expansive learning" (Engeström 1987) could be modelled in a meta-level architecture for data mining where changes in the event vocabulary itself would be the object of mining (see, e.g., work on ontology mining such as Buitelaar et al. 2005).

Whilst data mining has considerable potential for constructing process models, it is important to be aware of its limitations and constraints. Process mining using heuristics is subject to all that can go wrong with data mining (Han and Kamber 2001) and inductive approaches in general. This means that the quality of a model depends on the quality and representativeness of the data on which it has been constructed. In addition to this general concern, process models may overfit or underfit the data. For that reason, the construction process needs to keep humans in the loop to ensure the quality and reliability of the model built, through a meaningful choice, aggregation and representation of the data and a meaningful interpretation of the data mining results. An example of this is Tada-Ed (Merceron and Yacef 2005) which brings the power of data mining to the teacher in the classroom: computational power is coupled with the tacit and explicit knowledge of the teacher who is the only person who can truly perform meaningful analysis.

As communication (for learning) becomes increasingly mediated by technology and as computer memory is becoming cheaper, more data is recorded, at a finer level of granularity and analysis can be performed faster, it becomes increasingly possible to move from variable-based to event-based analysis. The discovery of regularities, relationships and trends in such huge amounts of data, either inductively or theory-guided, will require computational support. The confluence of developments in computer science with the needs in the learning sciences is a promising basis for exciting new possibilities in the future, adding to approaches as described here and conducted by others (e.g., Schümmer et al. 2005).

Acknowledgments This research has been supported by a Discovery Grant from the Australian Research Council.

References

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

Abbott, A. (1990). Conceptions of time and events in social science methods: Causal and narrative approaches. *Historical Methods, 23*, 140–150.

Abbott, A., & Hrycak, A. (1990). Measuring resemblance in sequence data: An optimal matching analysis of musicians' careers. *The American Journal of Sociology, 96*, 144–185.

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

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

Argote, L., & Ingram, P. (2000). Knowledge transfer: A basis for competitive advantage in firms. *Organizational Behavior and Human Decision Processes, 82*(1), 150–169.

Aristotle (1941). The basic works of Aristotle. In: R. McKeon (Ed.), 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.

Bales, R. F., & Strodtbeck, F. L. (1951). Phases in group problem solving. *Journal of Abnormal and Social Psychology, 46*, 485–495.

Buitelaar, P., Cimiano, P., & Magnini, B. (Eds.). (2005). *Ontology learning from text: Methods, evaluation and applications*. Amsterdam: IOS Press.

Cassandras, C. G. (1993). *Discrete event systems*. Homewood: Richard D. Irwin.

Dillenbourg, P., Baker, M., Blaye, A., & O'Malley, C. (1995). The Evolution of Research on Collaborative Learning. In P. Reimann & H. Spada (Eds.), *Learning in humans and machines: Towards an interdisciplinary learning science* (pp. 189–211). London: Elsevier.

Emirbayer, M., & Mische, A. (1998). What is agency? *The American Journal of Sociology, 103*(4), 962–1023.

Engeström, Y. (1987). *Learning by expanding: An activity-theoretical approach to developmental research*. Helsinki: Orienta-Konsultit Oy.

Engeström, Y. (1999). Activity theory and individual and social transformation. In Y. Engestroem, R. Miettinen, & R.-L. Punanmäki (Eds.), *Perspectives on activity theory* (pp. 19–38). Cambridge: Cambridge University Press.

Gersick, C. J. G. (1988). Time and transition in work teams: Toward a new model of group development. *Academy of Management Journal, 31*, 9–41.

Gill, A. (1962). *Introduction to the Theory of Finite-state Machines*. New York: McGraw-Hill.

Giudici, P., & Passerone, G. (2002). Data mining of association structures to model consumer behaviour. *Computational Statistics and Data Analysis, 38*(4), 533–541.

Gollwitzer, P. M. (1986). Action phases and mind-sets. In R. M. Sorrentino & E. T. Higgins (Eds.), *Handbook of motivation and cognition* (pp. 53–92). New York: Guilford.

Gouran, D. S., & Hirokawa, R. Y. (1996). Functional theory and comunication in decision-making and problem-solving groups: An expanded view. In R. Y. Hirokawa & M. S. Poole (Eds.), *Communication and group decision making* (pp. 55–80). Thousand Oaks: Sage.

Han, J., & Kamber, M. (2001). *Data mining: Concepts and techniques*. San Francisco: Morgan Kaufman.

Jermann, P., Soller, A., & Muehlenbrock, M. (2001). From mirroring to guiding: a review of the state of the art technology for supporting collaborative learning. In P. Dillenbourg, A. Eurelings, & K. Hakkarainen (Eds.), *European perspectives on computer-supported learning* (pp. 324–331). Maastricht: University of Maastricht.

Kapur, M., Hung, D., Jacobson, M. J., Voiklis, J., Kinzer, C. K., & Victor, C. D.-T. (2007). Emergence of learning in computer-supported, large-scale collective dynamics: A research agenda. Proceedings of the International Conference on Computer-supported Collaborative Learning (CSCL2007). New Brunswick, NJ.

Kaufman, L., & Rousseeuw, P. J. (1990). *Finding groups in data. An introduction to cluster analysis*. New York: Wiley.

Kay, J., Maisonneuve, N., Yacef, K., & Reimann, P. (2006). The Big Five and Visualisations for Team Work Activity. In M. Ikeda, K. D. Ashley, & T.-W. Chan (Eds.), *Proceedings of intelligent tutoring systems (ITS06)* (pp. 197–206). Heidelberg: Springer.

Kindler, E., Rubin, V., & Schäfer, W. (2006). *Process Mining and Petri Net Synthesis. Lecture Notes in Computer Science* (Vol. 4103, pp. 105–116). Business Process Management Workshops, Berlin: Springer.

McGrath, J. E., & Argote, L. (2001). Group processes in organizational contexts. In M. A. Hogg & R. S. Tindale (Eds.), *Blackwell handbook of social psychology* (Vol. 3, pp. 603–627). Oxford: Blackwell.

McGrath, J. E., & Beehr, T. A. (1990). Time and the stress process: Some temporal issues in the conceptualization and measurement of stress. *Stress Medicine, 6*, 95–104.

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

McIntyre, R. M., & Salas, E. (1995). Measuring and managing for team performance: Emerging principles from complex environments. In R. A. Guzzo & E. Salas (Eds.), *Team effectiveness and decision making in organizations* (pp. 9–45). San Francisco: Jossey-Bass.

Merceron, A., & Yacef, K. (2005). TADA-Ed for Educational Data Mining. *Interactive Multimedia Electronic Journal of Computer-Enhanced Learning*, *7*(1), http://imej.wfu.edu/articles/2005/2001/2003/index.asp.

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

Monge, P. R. (1990). Theoretical and analytical issues in studying organizational processes. *Organization Science, 1*(4), 406–430.

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.

Perera, D., Kay, J., Koprinska, I., Yacef, K., & Zaiane, O. (2008). Clustering and sequential pattern mining of online collaborative learning data. *IEEE Transactions on Knowledge and Data Engineering, 21*(6), 759–772.

Poole, M. S., & Doelger, J. A. (1986). Developmental processes in group decision-making. In R. Hirokawa & M. S. Poole (Eds.), *Communication and group decision-making* (pp. 35–62). Berverly Hills: Sage.

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

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.

Reimann, P. (2007). Time is precious: Why process analysis is essential for CSCL (and also can help to bridge between experimental and descriptive methods). In C. Chinn, G. Erkens, & S. Puntambekar (Eds.), *Minds, minds, and society. Proceedings of the Computer-supported Collaborative Learning Conference (CSCL 2007)* (pp. 598–607). New Brunswick: International Society of the Learning Sciences.

Reimann, P., Frerejean, J., & Thompson, K. (2009). Using process mining to identify models of group decision making processes in chat data. In C. O'Malley, D. Suthers, P. Reimann, & A. Dimitracopoulou (Eds.), *Computer-supported collaborative learning practives: CSCL2009 conference proceedings* (pp. 98–107). International Society for the Learning Sciences.

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

Salas, E., Sims, D. E., & Burke, C. S. (2005). Is there a "Big Five" in teamwork? *Small Group Research, 36*(5), 555–599.

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

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.

Searle, J. R., Kiefer, F., & Bierwisch, M. (1980). *Speech act theory and pragmatics*. Dordrecht: Kluwer Academic.

Sterman, J. D. (2000). *Business dynamics. Systems thinking and modeling for a complex world*. New York: McGraw-Hill.

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.

Tuckman, B. W. (1965). Developmental sequences in small groups. *Psychological Bulletin, 65*, 384–399.

Tuckman, B. W., & Jensen, M. A. C. (1977). Stages of small-group development revisited. *Group and Organizational Studies, 2*, 419–427.

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 Press.

Weijters, A. J. M. M., Aalst, W. M. P. V. D., & Medeiros, A. K. A. D. (2006). *Process mining with the heuristics miner-algorithm. BETA Working Paper Series WP 166*. Eindhoven: Eindhoven University of Technology.

Weinberger, A., & Fischer, F. (2006). A framework to analyze argumentative knowledge construction in computer-supported collaborative learning. *Computers & Education, 46*(1), 71–95.

Wheelan, S. A. (1994). *Group processes: A developmental perspective*. Sydney: Allyn & Bacon.

Zumbach, J., & Reimann, P. (2003). Influence of feedback on distributed problem based learning. Paper presented at the CSCL 2003 conference, June 15th to 18th, Bergen, Norway.

Chapter 9
Multilevel Analysis in CSCL Research

Jeroen Janssen, Gijsbert Erkens, Paul A. Kirschner, and Gellof Kanselaar

Abstract The aim of this chapter is to explain why multilevel analysis (MLA) is often necessary to correctly answer the questions CSCL researchers address. Although CSCL researchers continue to use statistical techniques such as analysis of variance or regression analysis, their datasets are often not suited for these techniques. The first reason is that CSCL research deals with individuals collaborating in groups, often creating *hierarchically nested datasets*. This means that such datasets for example contain variables measured at the level of the individual (e.g., learning performance) and variables measured at the level of the group (e.g., group composition or group performance). The number of unique observations at the lowest level, the individual, is higher than at the highest level, the group. Related to this, CSCL datasets often contain *differing units of analysis*. Some variables that CSCL researchers are interested in are measured at the individual level (e.g., gender, interactive behavior, familiarity with other group members), whereas other variables are measured at the group level (e.g., gender group composition, group performance, group consensus). Finally, because group members interact with each other in CSCL environments, this leads to *nonindependence* of the dependent variable(s) in the dataset. Because of their common experience during the collaboration, students' scores on the dependent variables will likely correlate (e.g., in a group with a relatively long history of successful collaboration, group members will report similar, high levels of trust, while in groups with a negative collaboration history, group members will likely report low levels of trust). Whether nonindependence is present in a dataset can be established by calculating the intraclass correlation coefficient. Whenever researchers encounter datasets with hierarchically nested data, differing units of analysis, and nonindependence, MLA is needed to appropriately model this data structure since it can appropriately disentangle the effects of the different levels on the dependent variable(s) of interest. Researchers however also employ other strategies to deal with nonindependence and hierarchy in their datasets (e.g., ignoring nonindependence and hierarchy, or

J. Janssen (✉)
Research Centre Learning in Interaction, Utrecht University, Utrecht, The Netherlands
e-mail: J.J.H.M.Janssen@uu.nl

S. Puntambekar et al. (eds.), *Analyzing Interactions in CSCL*, Computer-Supported Collaborative Learning Series, DOI 10.1007/978-1-4419-7710-6_9,

aggregating or disaggregating their data). We will highlight the dangers of these strategies using examples from our own research (e.g., increasing the chance of committing a Type I error). The chapter ends with a discussion of the advantages and disadvantages of using MLA for CSCL research. For example, although MLA is a powerful technique to address the data analytical problems CSCL researchers encounter, relatively large sample sizes are necessary.

9.1 Introduction

CSCL researchers are often interested in the processes that unfold between learners in online learning environments and the outcomes that stem from these interactions. However, studying collaborative learning processes is not an easy task. Researchers have to make quite a few methodological decisions such as how to study the collaborative process itself (e.g., develop a coding scheme or a questionnaire), on the appropriate unit of analysis (e.g., the individual or the group), and which statistical technique to use (e.g., descriptive statistics, analysis of variance, correlation analysis). Recently, several researchers have turned to multilevel analysis (MLA) to answer their research questions (e.g., Cress 2008; De Wever et al. 2007; Dewiyanti et al. 2007; Schellens et al. 2005; Strijbos et al. 2004; Stylianou-Georgiou et al. this volume). However, CSCL studies that apply MLA analysis still remain relatively scarce. Instead, many CSCL researchers continue to use ‘traditional’ statistical techniques (e.g., analysis of variance, regression analysis), although these techniques may not be appropriate for what is being studied. An important aim of this chapter is therefore to explain why MLA is often necessary to correctly answer the questions CSCL researchers address. Furthermore, we wish to highlight the consequences of failing to use MLA when this is called for, using data from our own studies.

9.2 Multilevel Analysis: A ‘New’ Methodological Approach in CSCL Research

Over the last 5 years or so, multilevel analysis (MLA) has been adopted by several CSCL researchers to answer their research questions, because MLA is especially suited to “appropriately grasp and disentangle the effects and dependencies on the individual level, the group level, and sometimes the classroom level” (Strijbos and Fischer 2007, p. 391). Although MLA is a relatively ‘new’ technique, especially to CSCL researchers, its development already began in the 1980’s (Snijders and Bosker 1999) and is since then used in several research disciplines.

MLA was initially embraced by educational researchers interested in school effectiveness research because it is well suited to the type of datasets they analyze. Consider, for example, an educational researcher interested in the effect of class size (i.e., the independent variable) on student achievement (i.e., the dependent variable). To investigate this effect, he or she would collect data about class sizes in different schools, as well as data on student achievement (e.g., standardized test scores on language, mathematics, and so on). However, such a research question poses several problems. First, this research question yields a *hierarchically nested dataset* with students nested within classrooms, and with classrooms nested within schools (and if this was an international study, even with schools nested within countries). Furthermore, the researcher would encounter the problem of *nonindependence* of his/her dependent variable of interest. Many statistical techniques (e.g., *t*-test, analysis of variance, regression analysis), assume that the achievement scores (or other dependent variables) of the students in the dataset are independent from each other (Hox 2003; Kashy and Kenny 2000; Snijders and Bosker 1999). In the example just provided, this will probably not be the case. Due to a common experience for example (e.g., the teaching they receive by their teacher), the scores within one classroom may not be independent at all since the overall classroom environment will affect all children in the class and even the behavior of individuals in the class will affect the others. Finally, the imaginary educational researcher would have to take into account that his or her variables of interest, class size and student achievement, are measured at different levels. Class size is measured at the class level, while achievement is measured at the student level. The number of available observations for both variables differs (i.e., the number is smaller for class size than for achievement). To properly address these issues, MLA was developed, and since then it became an important technique for school effectiveness research (Bosker and Snijders 1990; De Leeuw and Kreft 1986).

Social psychologists have also acknowledged the analytical problems described above. They are frequently interested in how individuals' thoughts and behaviors are influenced by other people. Many social psychological concepts involve two or more persons (e.g., attraction, interactive behavior, marital satisfaction) and thus the behavior of individuals *within a group* is often the focus of study (Kashy and Kenny 2000). A social psychologist might for example be interested in how the division of household chores (i.e., the independent variable) affects marital satisfaction (i.e., the dependent variable). To answer this question, the researcher would have to observe married couples and record who did which chore (and calculate for example a ratio), and to administer a questionnaire to both spouses to measure their marital satisfaction. From this example it becomes clear the social psychologist encounters the same problems as the educational researcher does. Both encounter the problem of hierarchically nested datasets (in this case individuals nested within couples), both involve variables at different levels of measurement (in this case the household chores ratio is measured at the level of the couple, whereas marital satisfaction is measured at the level of the individual), and in both cases the observations of the dependent variable are probably not independent (in this case there might even be a negative relationship: the husband may be more satisfied if he does little housework, while this may negatively affect the marital satisfaction of his wife).

The problem of statistical nonindependence of dependent variables (i.e., group members exerting a psychological influence on each other) has received considerable attention in social psychology since the 1980's (e.g., Bonito 2002; Kenny 1995, 1996; Kenny and Judd 1986, 1996) but less so in CSCL research (notable exceptions are for example, Cress 2008; Strijbos et al. 2004; Stylianou-Georgiou et al., see chapter #). It is therefore not surprising that social psychologists frequently use MLA to deal with these issues (cf., Bonito and Lambert 2005; Campbell and Kashy 2002; Kenny et al. 2002), while this technique is less often used in CSCL research.

9.3 The Problems CSCL Researchers Encounter

Similar to other research disciplines, CSCL researchers encounter the abovementioned problems of hierarchically nested datasets, nonindependence of dependent variables, and differing units of analysis. We explain these problems below.

9.3.1 Hierarchically Nested Datasets

In CSCL-environments, students work in groups. Studying online collaboration therefore often involves investigating group processes and how these processes are affected by contextual factors (e.g., the environment itself, the composition of the group, prior knowledge and experiences of the group members). It is not difficult to understand this leads to *hierarchically nested datasets*, since groups consist of two or more individuals and thus in these cases individuals are nested within groups. In many cases, CSCL researchers will encounter at least two levels: the group and the individual. The group is then the macro- or level-2 unit and the individual the micro- or level-1 unit (Hox 2003; Snijders and Bosker 1999). CSCL researchers may also use datasets that have even more levels of analysis. A researcher might for example be interested in the effects of the teacher's experience with CSCL on the way his or her students collaborate online. This researcher will have a dataset with three levels: students are nested within groups, while groups are nested within teachers' classrooms. Another CSCL researcher might be interested in the development of students' online interactive behavior over time. This researcher would therefore collect data about students' interactive behavior on different measurement occasions. This would also lead to a dataset with three levels: measurement occasions are nested within students, and students are nested within groups (Kenny et al. 2006; Snijders and Bosker 1999). Whenever researchers encounter datasets with hierarchically nested data, MLA is needed to appropriately model this data structure since it can appropriately disentangle the effects of the different levels on the dependent variable(s) of interest (Snijders and Bosker).

9.3.2 Nonindependence of Dependent Variables

Because their participants work in groups, CSCL researchers also encounter the problem of *nonindependence* of their dependent variables (Cress 2008). This means students within a group may be more similar to each other than are persons from different groups (Kenny et al. 2002). In the case of CSCL, the main source of this nonindependence is the mutual influence group members have on each other (Bonito 2002; Kenny 1996). In our own studies, which we describe in detail later on in this chapter, students could discuss with each other through a Chat-window and a Forum. Through these discussions, students influenced each other. In some cases for example, a student displayed negative behavior, and this prompted the other group members to respond negatively as well. Furthermore, some students in our studies were very active in the Chat conversations (e.g., they proposed a lot of strategies and asked a lot of questions). This could have triggered the other group members to also become more active in the chat as well. Such an influence of students on their group members' communication and behavior is nearly always present in CSCL research, because in CSCL-environments students communicate and collaborate to solve complex problems (Kreijns et al. 2003).

This reciprocal influence of group members is not necessarily positive, it can also be negative. In the previously mentioned example concerning active group members stimulating other group members to be more active, the reverse could also happen: When one group member is very active in the learning environment, this may trigger other group members to "sit back" and do little since that other group member is doing so much (O'Donnell and O'Kelly 1994; Webb and Palincsar 1996). Kenny et al. (2002) therefore noted that mutual influence can not only cause students to behave more similarly, but may also cause students to behave differently from their group members. This is called the *boomerang effect* (Kenny et al. 2006). Another example is that when group members behave negatively, a student may decide to counter this by displaying more positive behavior. Role assignment (cf., Schellens et al. 2005; Strijbos et al. 2004; Strijbos et al. 2007) may also lead to differential behavior. If one group member, for example, is given the task to ask critical questions, while the other group member has to monitor task progress, this may lead to differing behavior (e.g., the first student will ask many questions, but will display less metacognitive behavior, while the second student may display high levels of metacognitive behavior but may ask fewer questions). Kenny et al. therefore make a distinction between *positive nonindependence* where group members influence each other in such a way that they behave more similarly and *negative nonindependence* where group members influence each other to behave differently. Thus, since group members influence each other in a group context, this will likely lead to either positive or negative nonindependence of the dependent variables that are being investigated which in turn has to be dealt with during data analysis.

The degree of nonindependence can be estimated using the *intraclass correlation coefficient*[1] (ICC, cf., Kashy and Kenny 2000; Kenny et al. 2002). Values of the ICC can range from −1 to +1. An ICC of +1 for satisfaction with the collaborative process (scored on a 4-point scale ranging from 1 to 4) for example, indicates that when a group member has a score of 4 on this measure, the other group members will also have a

[1] For an excellent description on how to compute the ICC for a specific dataset, the reader is referred to Kenny et al. (2006).

score of 4. Conversely, an ICC of −1 for the same measure indicates that when one student has a score of 4 on this measure, his or her partners will have a score of 1.

An alternative interpretation of the ICC is in terms of the *amount of variance that is accounted for by the group* (Kenny et al. 2006). When the ICC for satisfaction with the collaborative process is found to be .40 for example, this means the 40% of the variance in this measure is accounted for by the group, and thus that 60% is accounted for by other (e.g., individual) factors.

The dependent measures that CSCL researchers are interested in will often be non-independent (Cress 2008). Strijbos et al. (2004) for example, studied the effect of roles on perceived group efficiency. They found an ICC of .47, meaning 47% of this measure is accounted for by the group. Group members displayed rather similar levels of perceived group efficiency, probably due to their common experiences in the CSCL environment. In a related study, Strijbos et al. (2007) found a similar influence of group level factors on group members' individual perceptions (ICC = .45). In the two studies described in the chapter by Stylianou-Georgiu et al. (chapter #), group level factors explained 11% and 8% of the variance respectively. On the other hand, not all researchers find similar substantial amounts of variance accounted for by the group. De Wever et al. (2007) for example, report only 3% of the students' level of knowledge construction was linked to the group level. However, these examples still illustrate the presence of nonindependence in datasets of CSCL researchers.

Nonindependence needs to be addressed when conducting statistical analyses, because it distorts estimates of error variances, thus making standard errors, *p*-values, and confidence intervals invalid when this distortion is not taken into account (Kenny 1995; Kenny et al. 2006). Traditional statistical techniques such as *t*-tests, analyses of variance, and regression analyses cannot cope with this distortion because they assume the variables are independent. Therefore CSCL researchers using these types of analyses run an increased risk of committing Type I or Type II errors (Kashy and Kenny 2000). Whether the chance to falsely reject (Type I error) or falsely accept (Type II error) the null hypothesis is increased, depends on the sign of the ICC (either positive or negative), and the type of dependent variable for which the ICC was calculated (see Kashy and Kenny for a detailed discussion).

Like any correlation coefficient, the ICC can be tested for significance. When the ICC is significant, its effect is large enough to bias statistical tests as described above (Kenny et al. 2006). However, because sample sizes are often small in CSCL research, the ICC may not be significant, while it is actually large enough to bias standard errors, *p*-values and so on. Kenny et al. (2002) therefore propose assuming group data are nonindependent even though the ICC is not significant.

9.3.3 Differing Units of Analysis

A final problem that CSCL researchers encounter concerns the *differing units of analysis* their datasets often contain. This has to do with the abovementioned hierarchical structure of their datasets. Some variables that CSCL researchers are interested in are measured at the individual level (e.g., gender, interactive behavior, familiarity with other group members), whereas other variables are measured at the group level (e.g., gender group composition, group performance, group consensus). Savicki and Kelley (2000) for example, studied the effect of gender and gender group composition (male-only, female-only, or mixed) on satisfaction with online collaboration. Their dependent variable was measured at the individual level (satisfaction), while their two independent variables were measured at the both the individual (gender) and group (gender group composition) level. Thus, their dataset contained variables with differing units of analysis.

Another example comes from a study conducted by Schellens et al. (2007). During their study, students collaborated in asynchronous discussion groups. They were interested in the impact of individual variables (e.g., gender, learning style) and group variables (active group versus relatively inactive group) on students' final exam scores. Their analyses therefore included independent variables measured at both the individual and the group level, while their dependent variable was measured at the level of the individual student.

To be able to cope with the different units of analysis encountered by Savicki and Kelley (2000) and Schellens et al. (2007), MLA is needed, because traditional statistical techniques cannot properly take these differing units of analysis properly into account (Hox 2003; Snijders and Bosker 1999).

9.4 Common Analysis Strategies

In this section we describe three strategies that researchers can use to deal with the data analytical problems described in the previous sections, namely ignoring nonindependence of dependent variables, aggregating or disaggregating data, and MLA.

9.4.1 Ignoring Nonindependence

A first strategy, and also still the most common practice during the analysis of group data (Kenny et al. 2002), is to ignore the hierarchical structure of the dataset, the nonindependence, and the differing units of analysis and perform statistical techniques such as *t*-tests or (M)ANOVA's (Cress 2008). As we discussed previously, this biases significance tests of inferential statistics (e.g., *t*- or *F*-values), sometimes making tests too liberal and, and at other times, too conservative.

Ignoring nonindependence is a frequently encountered approach in CSCL research. Francescato et al. (2006) for example, studied differences in students' evaluations of collaboration in online and face-to-face learning groups. Among other things, they investigated whether online learning groups perceived differing levels of social presence and satisfaction with the collaborative process than face-to-face groups. However, they found no differences between online and face-to-face groups using analyses of variance, although there was a tendency for online groups to be more satisfied with the collaborative process ($p = .17$). Because this study involves students working in groups, the evaluations of Francescato et al.'s students are most likely nonindependent. However, their analyses fail to take nonindependence into account, and thus the p-values reported by the authors might be biased. This could lead to a false acceptance of the null hypothesis (i.e., no differences between face-to-face and online learning groups). Using a more appropriate statistical technique, MLA, Francescato et al. might have been able to demonstrate significant differences between face-to-face and online learning groups.

Another example comes from the work of Guiller and Durndell (2007) who studied the effect of gender on students' linguistic behavior in online discussion groups. Guiller and Durndell studied whether male more absolute adverbials (i.e., strong assertions such as 'obviously') and imperatives (i.e., giving commands) than female students. In order to answer this question they coded students' messages and classified each message in terms of the linguistic behavior shown by the students. Guiller and Durndell then used χ^2-analyses to determine whether male and female students differed with respect to these behaviors. Although the authors found male students to use more absolute adverbials and imperatives, the corresponding χ^2-values were not significant. However, by using χ^2-analyses they too ignored the nonindependence of their dependent variables. Again, group members communicated and discussed with each other, so therefore they likely influenced each other. Using MLA, Guiller and Durndell, might have been able to detect statistically significant differences between male and female students on use of certain linguistic behaviors.

9.4.2 Aggregating or Disaggregating Data

Another strategy to deal with the problems described in the previous section is to *aggregate* individual data to the level of the group (Snijders and Bosker 1999). This involves summing the scores of the individual group members to create an aggregated group score.

This strategy is used in a study described by Van der Meijden and Veenman (2005) Van der Meijden and Veenman compared dyads using face-to-face (FTF, $N=20$) and computer-mediated communication (CMC, $N=22$) with respect to exchange of high-level elaboration (e.g., elaborate explanations or requests for help). Students' collaboration was coded using a coding scheme. However, the percentages of high-level elaboration "were calculated by summing the individual code frequencies" (p. 843) and dividing these by the total number of utterances. An independent samples t-test was then used to establish whether FTF and CMC conditions differed significantly with respect to high-level elaboration. High-level elaboration was thus treated as a group level variable. Such an analysis however, ignores the fact that high-level elaboration is in essence an individual level variable (although it may be affected by group level variables). Furthermore, by aggregating to the group level, this analysis uses fewer observations for high-level elaboration than are available. For this variable only 20 + 22 = 42 observations are used, while in effect there are 42 * 2 = 84 observations. Therefore Van der Meijden and Veenman run the risk of committing a Type II error. Fortunately, in their study the differences between FTF and CMC were large enough to detect a significant difference between FTF and CMC groups with respect to the percentage of high-level elaborations exchanged.

The reverse strategy can also be applied: treating group level data as if they were measured at the individual level. This is called *disaggregation*. Consider for example, the study by Savicki et al. (1996) about the effects of gender group composition on students' satisfaction with the collaborative process. Group composition was measured at the group level (all male, all female, or mixed groups), while satisfaction was measured at the individual level (students completed a questionnaire individually). In total, their sample consisted of 6 groups and 36 students. Savicki et al. conducted an analysis of variance to examine whether group composition affected satisfaction. However, this analysis does not take into account that group composition was measured at the group level. Thus, Savicki et al.'s analysis uses 36 observations for the group composition variable, while in fact there are only 6 observations for this variable. This led to an exaggeration of the actual sample size for this variable and increased the chance of committing a Type I error (Snijders and Bosker 1999).

9.4.3 Multilevel Analysis

MLA was designed specifically to cope with hierarchically nested data (Hox 2003; Snijders and Bosker 1999). Furthermore, it is a useful technique when researchers use datasets that have different units of analysis, such as group and individual level variables (Kenny et al. 2006). Finally, MLA can deal with the nonindependence of observations that results from the mutual influence group members have on each other (Snijders and Bosker).

At present, MLA is slowly finding its way to the CSCL research community: more and more CSCL researchers are using MLA to analyze their data. In the previously mentioned study of Strijbos et al. (2004), two conditions were present: a condition in which specific roles (e.g., project planner, editor) were assigned to students and a condition without role assignment. Thus, condition was a group level independent variable. Perceived group efficiency was measured using several questionnaires, and was therefore an individual level dependent variable. It is not difficult to see that in the Strijbos et al. study hierarchically nested data were collected since students were nested in groups. Furthermore, their study employed variables measured at different units of analysis. Finally, as we previously mentioned, nonindependence was present in their dataset, since they reported an ICC of .47 for perceived group efficiency. Strijbos et al. therefore constructed a ML model with perceived group efficiency as dependent variable and condition (role or non-role assignment) as an independent variable. Using MLA, they were able to model the nonindependence in their datasets and to analyze their dependent and independent variable at their appropriate levels of analysis.

9.5 Illustration of Problems and Analysis Strategies

In this section we will illustrate more elaborately the three problems (hierarchically nested datasets, nonindependence, and differing units of analysis) and strategies for data analysis (ignoring nonindependence, aggregating or disaggregating, and MLA) that were described in the previous sections. In order to do so, we utilize data from two different studies we conducted to illustrate three examples.

9.5.1 Example 1: Impact of an Awareness Tool on Online Discussion

The first example comes from the data collected for a study described in Janssen et al. (2007). For this study we developed an awareness tool (cf., Engelmann et al. 2009), called the Shared Space, which visualized the amount of agreement or discussion among group members during online synchronous chat discussions. We hypothesized that giving students such an awareness tool, would raise their awareness about the way they conducted their online discussions. In one condition students used the Shared Space (SS) to communicate online, while in the other condition (No SS) the students communicated through a regular chat-tool. We examined – amongst others – the effect of experimental condition on the number of times students evaluated the social interaction positively during their online conversations.

During this study we encountered the three abovementioned problems. First, because in this study students worked in groups, we had a *hierarchically nested dataset*. Furthermore, we found an ICC of .41 for our dependent variable, indicating a considerable influence of the group on this variable and the presence of *nonindependence*. In this study we also encountered the problem of *differing units of analysis*. Our dependent variable, the number of positive evaluations communicated by the students, was measured at the level of the individual. Because the group as a whole was assigned to either the SS or No SS condition, our independent variable, experimental condition, was measured at the level of the group. Moreover, we also wanted to control for students' level of participation, because some students were more active in the online discussions than others. Thus, our analysis also included a covariate, also measured at the level of the individual.

As we described before, we have three options when analyzing our data. If we chose the first option, *ignoring nonindependence*, we could use regression analysis to answer the question whether the Shared Space had an effect on the number of times students evaluated the collaboration positively. In this regression analysis, we include number of positive evaluations of the collaboration a student typed in the Chat-tool as a dependent variable and condition (effect coded with Shared Space as +1 and No Shared Space as −1) as an independent variable. Furthermore, we also include participation (e.g., the total number of messages students sent) in the regression equation to control for the fact that some students were more active during the online collaboration than others. As can be seen in Table 9.1, we find no effect of condition (Shared Space or No Shared Space) using this regression model on positive evaluations of the collaboration, $B=0.20$, $SE=0.14$, $p=.08$ (one-tailed significance). Thus, if we adopt a strategy that ignores nonindependence, we would conclude that the Shared Space does not influence the number of positive evaluations of their collaboration typed by students.

If we chose the second option, namely to *aggregate* our data, we could calculate the sum of positive evaluations of the collaboration for each group. On average, Shared Space groups exchanged 2.25 ($SD=3.16$) positive evaluations of the collaboration, while No Shared Space groups exchanged only 1.00 ($SD=1.95$) of these messages. Again, we could then use a regression analysis to examine the effects of condition (Shared Space or No Shared Space) on the number of positive evaluations of the collaborative process exchanged by the group. This regression analysis includes number of positive evaluations as the dependent variable, condition as the independent variable, and again level of participation (i.e., the total number of messages sent by the whole group) as a control measure. The results of the regression analysis are displayed in Table 9.2. As can be seen, condition was not found to have a significant impact on the number of positive evaluations of the collaborative process sent, $B=1.20$, $SE=0.83$, $p=.06$. This yields a conclusion comparable to the previously described strategy of ignoring nonindependence: the Shared Space does not have an influence on the amount of positive evaluations of the collaborative process exchanged during online collaboration.

Our final option is to use *multilevel analysis* to study the effects of the Shared Space on students' use of positive evaluations of the collaboration. In our study, we constructed a ML model that included number of times a student typed a positive evaluation of the collaboration as a dependent variable and condition (Shared Space

Table 9.1 Regression analysis of the effect of the shared space on number of positive evaluations of the collaborative process typed in the chat-tool

	B	SE B	β
Condition (−1 = No SS, +1 = SS)	0.202	0.142	.131
Participation	0.001	0.001	.140

Table 9.2 Regression analysis of the effect of the shared space on number of positive evaluations of the collaborative process typed in the chat-tool by the group

	B	SE B	β
Condition (−1 = No SS, +1 = SS)	1.202	0.830	.228
Participation	0.001	0.001	.166

Table 9.3 Multilevel analysis of the effect of condition (shared space or no shared space) on number of positive evaluations of the collaborative process exchanged

	β	SE
Participation	0.01*	0.00
Condition (SS or No SS)	0.20*	0.14
Deviance	429.14	
Decrease in deviance	2.06*	

*$p < .05$

Table 9.4 Summary of differing results for effects of shared space on students' positive evaluations of the collaborative process

	Ignoring nonindependence	Aggregating data	Multilevel analysis
Statistical analysis	Regression analysis	Regression analysis	MLA
Significance of effect of condition	Not significant, p = .08	Not significant, p = .08	Significant, p = .04
Conclusion	No effect of shared space on positive evaluations of collaborative process	No effect of shared space on positive evaluations of collaborative process	Positive effect of shared space on positive evaluations of collaborative process

or No Shared Space) as an independent variable. Furthermore, we included participation (e.g., total number of messages sent) again to control for the fact that some students typed more messages than other students. As can be seen in Table 9.3, we found a significant effect of the Shared Space on the number of positive evaluations of the collaboration students typed, β=0.20, SE = .14, p = .04 (one-tailed significance). Although the differences in p-values are small (see Table 9.4) and the differences may not seem spectacular, this last analysis strategy leads to a different conclusion than ignoring nonindependence or aggregating data, namely that the Shared Space affects the number of positive evaluations of the collaboration. Thus, in this case MLA prevented us from making a Type II error (i.e., falsely accepting the null hypothesis).

9.5.2 Example 2: Influence of Representational Guidance on Student Learning

Our second example comes from a study reported in Janssen et al. (2010). In this study we investigated the effects of representational guidance (cf., Suthers 2001; Suthers and Hundhausen 2003) on students' performance on a knowledge post-test. Our design used two conditions: In one condition students used a Graphical

Debate-tool to construct external representations of a historical debate, while in the other condition students used a Textual Debate-tool to construct such representations. Both versions of the tool differed with respect to the representational guidance they offered to the students. The Graphical Debate-tool made extensive use of visualization techniques to visualize certain aspects of the collaborative problem solving process (e.g., Was there a balance between the number of arguments pertaining to both positions?). We hypothesized that the representational guidance offered by the Graphical Debate-tool would positively affect students' post-test performance.

Again we encountered the previously mentioned problems during our study. In this study too, students worked in groups, which created a *hierarchically nested dataset*. When we calculated the ICC of our dependent variable, post-test performance, we found an ICC of .32. This meant that 32% of the total variance was explained by group level variables and that the assumption of *independence* was violated. Finally, the variables we studied were measured at *different units of analysis*. Post-test performance, our dependent variable, was measured at the level of the student. In contrast, our independent variable, experimental condition (Graphical versus Textual Debate-tool) was measured at the level of the group, because each group was assigned to one of the two conditions. Finally, our analyses also included a covariate, pretest performance, which was again measured at the individual level.

If we chose to *ignore nonindependence* when analyzing the effects of the Graphical Debate-tool on students' post-test performance, we could use analysis of covariance (ANCOVA). The ANCOVA model would include post-test performance as the dependent measure of interest, condition (Graphical versus Textual Debate-tool) as the independent variable, and pre-test performance as a covariate. The results of this analysis can be found in Table 9.5. As can be seen in this Table, condition had a significant impact on post-test performance, $F(1, 82) = 3.98$, $p = .05$. In conclusion, if we adopt a strategy that ignores nonindependence, we would conclude that condition has a significant impact on post-test performance.

Our second option would be to *aggregate* our data. This involves computing, for each group, the average post- and pre-test score of the individual group members. Using such a strategy, we find Graphical Debate groups to attain, on average, a post-test score of 13.02, while Textual Debate groups attain a an average score of 12.24. To test the effect of condition on post-test performance, we could again conduct an analysis of covariance, using post-test performance as the dependent variable, condition as the independent variable, and pre-test performance as a covariate. The results of this analysis are displayed in Table 9.6. As can be seen, the

Table 9.5 Analysis of covariance for condition (graphical versus textual debate-tool) on post-test performance

	df	MS	F	η^2
Pretest performance (covariate)	1	39.14	11.11**	.12
Condition (graphical or textual debate)	1	14.07	3.98*	.05
Error	82	3.55		

$^{*}p < .05$; $^{**}p < .01$

Table 9.6 Analysis of covariance for condition (graphical versus textual debate-tool) on group level variables

	df	MS	F	η^2
Pretest performance (covariate)	1	11.39	6.27*	.28
Condition (graphical or textual debate)	1	0.50	0.28	.02
Error	37	1.82		

*$p < .01$

Table 9.7 Multilevel analysis of the effect of condition (graphical or textual debate-tool) on post-test performance

	β	SE
Pre-test performance	0.28**	0.10
Condition (graphical or textual)	0.42*	0.22
Deviance	344.63	
Decrease in deviance	11.07**	

*$p < .05$; **$p < .01$

Table 9.8 Summary of differing results for effects of graphical debate-tool on students' post-test performance

	Ignoring nonindependence	Aggregating data	Multilevel analysis
Statistical analysis	Analysis of covariance	Analysis of covariance	MLA
Significance of effect of condition	Significant, p = .05	Not significant, p = .61	Significant, p = .03
Conclusion	Positive effect of graphical debate-tool on post-test performance	No effect of graphical debate-tool on post-test performance	Positive effect of graphical debate-tool on post-test performance

effects of condition are not significant if we adopt an aggregation strategy, $F(1, 37) = 0.28$, $p = .61$. This means we would conclude, in contrast to the previous strategy of ignoring nonindependence, the Graphical Debate-tool does not have a positive effect on students' post-test performance.

Our final option is to conduct a *multilevel* analysis. Our ML model then includes students' pre-test performance and condition (Graphical versus Textual Debate-tool). The results of this analysis can be found in Table 9.7. We found a significant effect of condition on post-test performance, indicating that the Graphical Debate-tool helped students to perform better on the knowledge post-test, $\beta = 0.42$, SE = .22, $p = .03$.

Table 9.8 summarizes the results of the different analysis strategies. As can be seen, the p-values are different if one strategy is chosen rather than another strategy. Especially when an aggregation strategy is chosen for the evaluation of the effect of the Graphical Debate-tool, a difference is noticeable. When we analyzed aggregated data we would draw a different conclusion – the Graphical Debate-tool does not affect post-test performance – compared to ignoring nonindependence or using MLA. This again highlights the importance of carefully using the appropriate data analysis strategy.

9.5.3 Example 3: Influence of Representational Guidance on Essay Quality

The study described in the previous example also provides the opportunity to highlight the effects of using a *disaggregation* strategy. Besides post-test performance, we also examined the effects of the Graphical Debate-tool on the quality of the essays written by groups. This effect was examined by measuring the number of topics covered in the essay and the quality of the essay. Because the essays were written by groups, this variable was a group-level measure: each group received one score for number of topics covered and quality of the essay. Thus, multilevel analysis is not necessary. However, in this could we could also have adopted a disaggregation strategy. This means each student within the group is given the same score for these two quality indicators. This leads to an increase of the sample size from 39 groups to 124 students. In Janssen et al. (2010), using t-tests, we found no significant differences with respect to the number of topics covered, $t = -0.55$, $p = .59$, and quality of the essay, $t = 2.00$, $p = .06$. However, when we disaggregate our data, and then use t-tests to examine the differences between the Graphical and Textual Debate tool, we find different t- and p-values, namely for number of topics covered, $t = -1.38$, $p = .17$, and for essay quality, $t = 3.24$, $p = .00$. This example shows that using a disaggregation strategy might lead to biased t- and p-values and even different conclusions (i.e., in the case of essay quality the conclusion would be different).

9.6 Conclusion and Discussion

In this chapter we discussed the data analytical problems CSCL researchers frequently encounter, namely hierarchically nested datasets, nonindependence of dependent variables, and differing units of analysis. We argued that, in order to take these problems into account, MLA should be used. We also demonstrated that alternative analysis strategies such as ignoring nonindependence or aggregating or disaggregating data can lead to different results and possibly to mistakes regarding the significance or non-significance of these results. We therefore strongly advocate the use of MLA in CSCL research. Fortunately, more and more CSCL researchers are beginning to use this technique to answer their research questions.

It should be noted that we do not claim that in the cases where CSCL researchers used other analyses than MLA their conclusions are wrong. This need not be the case. However, these researchers do have an increased chance of committing Type I or Type II errors. We hope this chapter will contribute to an increased awareness of the risks of using traditional statistical techniques such as *t*-tests and ANOVAs, and future CSCL research will use MLA when this is appropriate.

Of course not all data-analytic problems that CSCL researchers encounter are solved by using MLA. Furthermore, MLA has its own limitations. First, MLA is mostly used when the dependent variable is measured at the interval level of measurement. Sometimes however, researchers may be interested in dichotomous (e.g., success or failure of group work) or categorical dependent variables (e.g., levels of knowledge construction). Although MLA techniques have been developed to incorporate these kinds of dependent variables (multilevel logistic regression, see Snijders and Bosker 1999), they are rarely adapted to CSCL data.

Second, for an adequate analysis of collaborative learning using MLA, it is often suggested that large sample sizes at all levels (individual as well as group) are necessary (Cress 2008; Maas and Hox 2005). Maas and Hox, using a simulation study, demonstrated that only a small sample size at the group level (less than 50 groups) is problematic and leads to biased estimates. A small sample size at the individual level (groups consisting of five group members or less), does not appear to be problematic. This means that, in order to use MLA confidently for CSCL data, researchers should collect data about at least 50 groups. CSCL researchers often employ less than 50 groups in their studies. Given the complexity of CSCL research and how time-consuming data collection and analysis often are, a sample size of at least 50 groups places a heavy burden on CSCL researchers.

Third, CSCL researchers are often interested in data over time. An example might be how familiarity with group members affects trust-development in CSCL environments over time. To investigate this question a researcher would collect data about trust levels on different occasions. This adds even more problems to analyzing CSCL data. The effects of familiarity on trust may not be the same at every measurement occasion (e.g., its effects may be greater at the beginning of the collaboration). Furthermore, the level of trust at measurement occasion 1 may also have an effect on the level of trust at occasion 2 (if trust was high at occasion 1, this may affect trust at occasion 2). This creates a new type of nonindependence: autocorrelation (Kenny et al. 2006). Again, MLA techniques have been developed to analyze time-series data (cf., Chiu and Khoo 2003, 2005; Kenny et al. 2006, but they are not often used in CSCL research. CSCL researchers should therefore begin to investigate the possibilities of using MLA for time-series data.

Finally, MLA will not be a suitable technique to answer all research questions. Quite a lot CSCL research focuses on capturing the interactive processes that unfold between group members. In some cases researchers are interested in providing "thick" or "rich" descriptions of the collaborative process (Baker 2003; Hmelo-Silver and Bromme 2007). In such cases, MLA is obviously useless. Furthermore, it has been argued that studying intersubjective meaning making or group cognition should be the focus of CSCL research (Stahl 2006; Suthers 2006). This involves studying "how people make sense of situations and of each other" (Suthers, p. 321).

Researchers with such a perspective on CSCL research could object to disentangling group and individual aspects of collaborative learning. They would argue that in order to understand the collaborative process, the group should be the unit of analysis, not the individual. Again, if one has such an approach to studying CSCL, using MLA will not be a sensible strategy.

Fortunately, over the last years a large body of literature on MLA has been published. For the CSCL researcher who finds him- or herself faced with a hierarchically nested dataset and nonindependent observations, several good and accessible textbooks on the statistical and technical background of MLA are available (e.g., Hox 2003; Snijders and Bosker 1999). Furthermore, several good articles have been published about how to apply MLA to group and CSCL data (e.g., Bonito 2002; Cress 2008; Kenny et al. 2006; Kenny et al. 2002) and several CSCL articles have been published that can serve as an example (De Wever et al. (2006); De Wever et al. 2007; Schellens et al. 2005; Schellens et al. 2007; Strijbos et al. 2004, 2007). Finally, several programs specifically designed for performing MLA are available, such as MLwiN (http://www.cmm.bristol.ac.uk/MLwiN/index.shtml) and HLM 6 (http://www.ssicentral.com/hlm/index.html). Moreover, the fact that conventional statistical software such as SPSS and SAS now incorporate procedures for carrying out MLA means that the possibility to perform MLA has become a possibility for many CSCL researchers.

CSCL research can still make progress by incorporating MLA in its repertoire of analysis techniques. It is an encouraging development that CSCL researchers are turning toward MLA more often. It is our hope and expectation that this development will continue and that CSCL researchers are going to find new ways to deal with the complex data analytical problems they are faced with. Ultimately, this will lead to a better understanding of how critical features of the CSCL-environment (e.g., support given by the environment), the group (e.g., composition), and the individual student (e.g., prior knowledge, motivation) affect social interaction and students' learning processes. Furthermore, when researchers combine MLA with qualitative analyses in a mixed methods design (Leech and Onwuegbuzie 2009) an even more complete picture of the CSCL process is possible.

References

Baker, M. (2003). Computer-mediated argumentative interactions for the co-elaboration of scientific notions. In J. Andriessen, M. Baker, & D. Suthers (Eds.), *Arguing to learn: Confronting cognitions in computer-supported collaborative learning environments* (pp. 47–78). Dordrecht: Kluwer Academic.

Bonito, J. A. (2002). The analysis of participation in small groups: Methodological and conceptual issues related to interdependence. *Small Group Research, 33*, 412–438.

Bonito, J. A., & Lambert, B. L. (2005). Information similarity as a moderator of the effect of gender on participation in small groups: A multilevel analysis. *Small Group Research, 36*, 139–165.

Bosker, R. J., & Snijders, T. A. (1990). Statistische aspecten van multi-niveau onderzoek (Statistical aspects of multilevel research). *Tijdschrift voor Onderwijsresearch, 15*, 317–329.

Campbell, L., & Kashy, D. A. (2002). Estimating actor, partner, and interaction effects for dyadic data using PROC MIXED and HLM: A user-friendly guide. *Personal Relationships, 9*, 327–342.

Chiu, M. M., & Khoo, L. (2003). Rudeness and status effects during group problem solving: Do they bias evaluations and reduce the likelihood of correct solutions? *Journal of Education and Psychology, 95*, 506–523.

Chiu, M. M., & Khoo, L. (2005). A new method for analyzing sequential processes: Dynamic multilevel analysis. *Small Group Research, 36*(5), 600–631.

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.

De Leeuw, J., & Kreft, I. (1986). Random coefficient models for multilevel analysis. *Journal of Educational Statistics, 11*, 57–85.

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

De Wever, B., Van Keer, H., Schellens, T., & Valcke, M. (2007). Applying multilevel modelling to content analysis data: Methodological issues in the study of role assignment in asynchronous discussion groups. *Learning and Instruction, 17*, 436–447.

Dewiyanti, S., Brand-Gruwel, S., Jochems, W., & Broers, N. J. (2007). Students' experiences with collaborative learning in asynchronous computer-supported collaborative learning environments. *Computers in Human Behavior, 23*, 496–514.

Engelmann, T., Dehler, J., Bodemer, D., & Buder, J. (2009). Knowledge awareness in CSCL: A psychological perspective. *Computers in Human Behavior, 25*(4), 949–960.

Francescato, D., Porcelli, R., Mebane, M., Cuddetta, M., Klobas, J., & Renzi, P. (2006). Evaluation of the efficacy of collaborative learning in face-to-face and computer-supported university contexts. *Computers in Human Behavior, 22*, 163.

Guiller, J., & Durndell, A. (2007). Students' linguistic behaviour in online discussion groups: Does gender matter? *Computers in Human Behavior, 23*, 2240–2255.

Hmelo-Silver, C. E., & Bromme, R. (2007). Coding discussions and discussing coding: Research on collaborative learning in computer-supported environments. *Learning and Instruction, 17*, 460–464.

Hox, J. (2003). *Multilevel analysis: Techniques and applications*. Mahwah: Lawrence Erlbaum.

Janssen, J., Erkens, G., & Kanselaar, G. (2007). Visualization of agreement and discussion processes during computer-supported collaborative learning. *Computers in Human Behavior, 23*, 1105–1125.

Janssen, J., Erkens, G., Kirschner, P. A., & Kanselaar, G. (2010). Effects of representational guidance during computer-supported collaborative learning. *Instructional Science, 38*, 59–88.

Kashy, D. A., & Kenny, D. A. (2000). The analysis of data from dyads and groups. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (pp. 451–477). Cambridge: Cambridge University Press.

Kenny, D. A. (1995). The effect of nonindependence on significance testing in dyadic research. *Personal Relationships, 2*, 67–75.

Kenny, D. A. (1996). Models of non-independence in dyadic research. *Journal of Social and Personal Relationships, 13*, 279–294.

Kenny, D. A., & Judd, C. M. (1986). Consequences of violating the independence assumption in analysis of variance. *Psychological Bulletin, 99*, 422.

Kenny, D. A., & Judd, C. M. (1996). A general procedure for the estimation of interdependence. *Psychological Bulletin, 119*, 138–148.

Kenny, D. A., Kashy, D. A., & Cook, W. L. (2006). *Dyadic data analysis*. New York/London: Guilford.

Kenny, D. A., Mannetti, L., Pierro, A., Livi, S., & Kashy, D. A. (2002). The statistical analysis of data from small groups. *Journal of Personality and Social Psychology, 83*, 126–137.

Kreijns, K., Kirschner, P. A., & Jochems, W. (2003). Identifying the pitfalls for social interaction in computer-supported collaborative learning environments: A review of the research. *Computers in Human Behavior, 19*, 335–353.

Leech, N. L., & Onwuegbuzie, A. J. (2009). A typology of mixed methods research designs. *Quality and Quantity, 43*, 265–275.

Maas, C. J. M., & Hox, J. J. (2005). Sufficient sample sizes for multilevel modelling. *European Journal of Research Methods for the Behavioral and Social Sciences, 1*, 85–91.

O'Donnell, A. M., & O'Kelly, J. (1994). Learning from peers: Beyond the rhetoric of positive results. *Educational Psychology Review, 6*, 321–349.

Savicki, V., & Kelley, M. (2000). Computer mediated communication: Gender and group composition. *Cyberpsychology and Behavior, 3*, 817–826.

Savicki, V., Kelley, M., & Lingenfelter, D. (1996). Gender and group composition in small task groups using computer-mediated communication. *Computers in Human Behavior, 12*, 209–224.

Schellens, T., Van Keer, H., & Valcke, M. (2005). The impact of role assignment on knowledge construction in asynchronous discussion groups: A multilevel analysis. *Small Group Research, 36*, 704–745.

Schellens, T., Van Keer, H., Valcke, M., & De Wever, B. (2007). Learning in asynchronous discussion groups: A multilevel approach to study the influence of student, group and task characteristics. *Behaviour and Information Technology, 26*, 55–71.

Snijders, T. A. B., & Bosker, R. J. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. London: Sage.

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

Strijbos, J.-W., & Fischer, F. (2007). Methodological challenges for collaborative learning research. *Learning and Instruction, 17*, 389–393.

Strijbos, J. W., Martens, R. L., Jochems, W. M. G., & Broers, N. J. (2004). The effect of functional roles on group efficiency: Using multilevel modeling and content analysis to investigate computer-supported collaboration in small groups. *Small Group Research, 35*, 195–229.

Strijbos, J. W., Martens, R. L., Jochems, W. M. G., & Broers, N. J. (2007). The effect of functional roles on perceived group efficiency during computer-supported collaborative learning: A matter of triangulation. *Computers in Human Behavior, 23*, 353–380.

Stylianou-Georgiou, A., Papanastasiou, E., & Puntambekar, S. (this volume). Analyzing collaborative processes and learning from hypertext through hierarchical linear modelling. In S. Puntambekar, G. Erkens, & C. Hmelo-Silver (Eds.), *Analyzing interactions in CSCL: Methods, approaches and issues*.

Suthers, D. D. (2001). Towards a systematic study of representational guidance for collaborative learning discourse. *Journal of Universal Computer Science, 7*, 254–277.

Suthers, D. D. (2006). Technology affordances for intersubjective meaning making. *International Journal of Computer Supported Collaborative Learning, 1*, 315–337.

Suthers, D. D., & Hundhausen, C. D. (2003). An experimental study of the effects of representational guidance on collaborative learning processes. *Journal of the Learning Sciences, 12*, 183–218.

Van der Meijden, H., & Veenman, S. (2005). Face-to-face versus computer-mediated communication in a primary school setting. *Computers in Human Behavior, 21*, 831–859.

Webb, N. M., & Palincsar, A. S. (1996). Group processes in the classroom. In D. C. Berliner (Ed.), *Handbook of educational psychology* (pp. 841–873). New York: Simon & Schuster Macmillan.

Chapter 10
Sequential Analysis of Scientific Argumentation in Asynchronous Online Discussion Environments*

Allan Jeong, Douglas B. Clark, Victor D. Sampson, and Muhsin Menekse

Abstract The national science standards call for increased focus on scientific argumentation in the classroom, and researchers have developed sophisticated online science learning environments to promote and support student engagement in scientific argumentation. Assessing the quality of scientific dialogic argumentation in these environments, however, has proven to be challenging. Existing analytic frameworks tend to assess scientific argumentation using the presence or absence of various types of comments (e.g., frequency of claims, rebuttals, and supporting statements) that do not fully convey the dynamic and dialogic nature of argumentation. In this chapter, we present a sequential analysis approach developed by Jeong (2005) that incorporates a coding scheme developed by Clark and Sampson (2007, 2008) to identify, visualize, and assess the dialogic processes of argumentation in online science learning environments in terms of transitional probabilities, transitional state diagrams, and other related measures. These measures include: (a) how and how often students respond to particular discourse moves (e.g., the probabilities that responses to claims are rebuttals *vs.* simple agreement *vs.* no response); and (b) how and to what extent observed response patterns produce extended chains of discourse moves that exhibit high levels of argumentation (claim → challenge → explain or amend claim). A sample analysis is presented to illustrate how this approach can also be used to assess how characteristics of the discourse environment affect the quality of argumentation and better understand the interplay between discourse environments and collaborative discourse.

---

*This work was partly funded by the National Science Foundation, grant 0334199, as part of the center for Technology Enhanced Learning in Science (TELS). Any opinions, findings, and conclusions or recommendations expressed in this study are those of the researchers and do not necessarily reflect the views of the National Science Foundation.

A. Jeong (✉)
Associate Professor, Florida State University, College of Education, Department of Educational Psychology & Learning Systems, Instructional Systems Program
e-mail: ajeong@fsu.edu

S. Puntambekar et al. (eds.), *Analyzing Interactions in CSCL*, Computer-Supported Collaborative Learning Series, DOI 10.1007/978-1-4419-7710-6_10,

10.1 Introduction

Online learning environments designed to engage and support students in dialogic scientific argumentation provide excellent opportunities for students to propose, support, evaluate, critique, and refine ideas in a more productive manner. Over the last decade, a number of sophisticated environments have been developed to support students engaging in this type of knowledge-building or knowledge-validating discourse. Examples, among others, include *CONNECT* (e.g., deVries et al. 2002), TC3 (e.g., Erkens et al. 2003), *DUNES* (e.g., Schwarz and Glassner in press), Virtual Collaborative Research Institute (e.g., Janssen et al. 2007), *ArgueGraph* (e.g., Jermann and Dillenbourg 2003), and the *personally-seeded discussions* within the Web-based Inquiry Science Environment (e.g., Clark 2004; Clark and Sampson 2007, 2008; Clark et al. 2009; Cuthbert et al. 2002). The multitude of approaches used to foster argumentation gives rise to complex and diverse assessment needs among researchers and an increasing interest in approaches for analyzing and assessing the nature or quality of dialogic scientific argumentation. To date, researchers have developed a broad range of methods that reflect various perspectives on argumentation, pedagogical goals, and curricular structures (see Clark et al. 2007 for a catalog of several of these methods). These methods tend to do an excellent job of providing overall ratings and observed frequencies of argumentative and collaborative interactions. However these frameworks tend not to provide information about the specific sequences of discourse moves produced in student exchanges – information that is needed to fully capture and computationally model the dynamic nature of argumentative discourse in CSCL (Jeong 2005).

For example, content analysis is one of the most common methods used in CSCL when analyzing learner interactions. In this method, researchers identify message categories and measure the frequency of messages observed in each category (Rourke et al. 2001). This approach generates results that are mainly descriptive rather than prescriptive in nature, reporting for example the frequencies of arguments, challenges, and explanations observed in a discussion. However, message frequencies provide little information that can be used to explain or predict how participants respond to given types of messages (e.g., argument → challenge *versus* argument → simple agreement), how response patterns are influenced by latent variables (e.g., message function, content, communication style, response latency) or exogenous variables (e.g., gender, personality traits, discussion protocols, type of task), and how particular response patterns contribute to observed differences in group performance on a desired outcome. Therefore, new approaches are needed to examine to what extent messages elicit responses based on what is said in conjunction with when, how, who, and why messages are presented, and whether or not the elicited responses help produce sequences of speech acts that support critical discourse (e.g., claim → challenge → explain) and group performance in decision making, problem-solving, and learning.

In this chapter, we integrate two complimentary methods that researchers can use in tandem to analyze and assess the nature of the interactions that take place between students in CSCL environments that use asynchronous threaded discussion forums to engage students in scientific argumentation. The first method, developed by Clark and Sampson (2007, 2008), codes the nature of the discourse moves, the quality of the grounds used to support and challenge ideas, and the level of opposition that takes place between students as they propose, support, critique and refine ideas. The second method, sequential analysis (Bakeman and Gottman 1997) and the tools used to perform this types of analysis (Jeong 2005), captures the dynamic and dialogic nature of argumentation by measuring how and how likely students respond to various discourse moves of interest (e.g., the probabilities that responses to claims are rebuttals *vs*. simple agreement *vs*. no response), and how and to what extent these observed response patterns produce extended chains of discourse moves that reveal processes essential to producing high quality argumentation (claim → challenge → explain or amend claim). In the sections that follow, we will: (a) outline the sequence of steps, tools, and metrics used in each approach; (b) conduct a sample analysis that illustrates how these two methods can be used in tandem to compare and contrast various aspects of scientific argumentation; and (c) discusses implications and recommendations for researchers interested in using these approaches in tandem.

10.2 Steps, Tools, and Metrics Used in Each Approach

This section describes a procedure for coding the nature of the contributions made by the participants in an asynchronous discussion forum and the oppositional level of various discourse episodes (method 1) and the steps used to perform a sequential analysis of the argumentative discourse (method 2). This first method (Clark and Sampson 2007, 2008) consists of four major steps: (a) coding the discourse moves observed in individual postings/comments; (b) coding the grounds of a comment; (c) parsing the discussions into discourse episodes; and (d) scoring the level of opposition found within discourse episodes. This type of analysis enables a researcher to focus on specific episodes found within a discussion and provides a way to document the extent to which students question or challenge each other's ideas, how often they use grounds to support or challenge an idea, and the conceptual quality of students' ideas. Once this analysis is complete, a researcher can use the second method, sequential analysis, to identify response patterns measured in terms of the probabilities in which certain types of responses are elicited by given types of comments. This particular method, developed and refined by Jeong (2004, 2005), enables researchers to look at the discussion forum as a whole (across multiple episodes) to compare and identify similarities and differences in patterns of discourse produced between different groups under different conditions – patterns that might help to explain the observed number of times students question and challenge one another's ideas, use grounds to support or challenge ideas, and the quality of students' ideas.

10.2.1 Core Coding: Examining the Nature of Comments Found within Discourse Episodes

10.2.1.1 Coding the Discourse Moves of Individual Postings

The framework assigns a discourse move code to each comment based on the comment's role in the discussion. In order to avoid ambiguity in terms of references within a comment, the framework codes each comment in relation to the parent comment to which it responds. These codes take into account comments that are typically examined as part of a structural analysis (e.g., claims, counter-claims, rebuttals), meta-organizational comments that help organize the interaction (which are typically overlooked in a structural analysis), and the occasional off-task interaction. The full list of *discourse move* comment codes is outlined in Table 10.1.

Table 10.1 Coding scheme for the discourse move of individual comments

Discourse move	Definition
Claim	The seed-comment principle or an assertion made.
Counter-claim	An assertion made by a pair of students that is different from (and does not attack) the seed claim or parent comment made by another pair of students. This code is only assigned when a comment does not focus on any aspect of the thesis of the comment it replies to; instead it offers an entirely new interpretation of the phenomena.
Change of claim	A comment made by a pair of students that indicates that: (1) they have changed their original claim; or (2) changed their viewpoint; or (3) have made a concession in response to comments (claims or rebuttals) made by another pair of students.
Rebuttal against grounds	An attack on, or disagreement with, the grounds (evidence, explanations, qualifiers, or backing) used by another pair of students to support or justify their comment.
Rebuttal against thesis	An attack on or disagreement with the thesis (or a specific part of the thesis) of another pair of students' comment (claim or rebuttal) that does not attack the grounds.
Clarification in response to a rebuttal	This code is assigned to comments used to strengthen a position (in terms of accuracy or validity) in response to a rebuttal without attacking the rebuttal or grounds made by another pair of students.
Support of a comment	A statement used to support the truth or accuracy of the previous claim or rebuttal. This category includes statements that: (1) voice agreement with a comment; (2) rewords the previous comment; (3) adds additional grounds in support; or (4) expands on the comment.

(continued)

Table 10.1 (continued)

Discourse move	Definition
Query about meaning	A comment that asks for clarification of an earlier comment (e.g., "What do you mean when you say...?" or "I don't understand what you are saying?"). These comments question the meaning of a statement rather than the accuracy of the statement.
Clarification of meaning	A comment made by a pair of students to clarify (restate in a new way) a previous comment. The purpose of these comments is to clarify the meaning of a statement in response to a query (about meaning) rather than supporting the accuracy of a statement
Organization of participants	A comment that: (1) reminds other participants to participate; (2) asks others for feedback; (3) has a meta-organizational aspect (e.g. "Do we all agree?"); (4) attempts to change the way someone else in the discussion is participating.
Off-task	Comments that are not about the topic (e.g., "Nice haircut, John!").

10.2.1.2 Coding the Grounds of a Comment

Rather than simply identifying the presence or absence of grounds, the framework classifies a comment as having no grounds (grounds quality level 0), including only an explanation without evidence as grounds (grounds quality level 1), using evidence as grounds (grounds quality level 2), and including evidence and an explanation or coordinating multiple pieces of evidence or as grounds (grounds quality level 3). We developed a series of binary decisions (see Fig. 10.1 flow chart) to increase reliability in the coding process. Whereas all comments receive a discourse move code, not all comments receive a grounds quality and conceptual quality code because for some comments (such as "organization of participation," "query about meaning," and "off-task"), these qualities simply to do not apply. Coding of grounds is not the focus of the current chapter, but full detail about this aspect of the coding scheme is available in Clark and Sampson (2008).

10.2.1.3 Coding the Conceptual Quality of a Comment

Finally, the *conceptual quality* of the comment is rated as either non-normative (*conceptual quality level 0*), transitional (*conceptual quality level 1*), normative (*conceptual quality level 2*), or nuanced (*conceptual quality level 3*). In coding a comment, the framework first determines how many non-normative, transitional, and normative facets are included as part of the entire comment using conceptual facet tables developed through extensive prior conceptual change work measuring the longitudinal evolution of students' conceptual ecologies (Clark 2000, 2006;

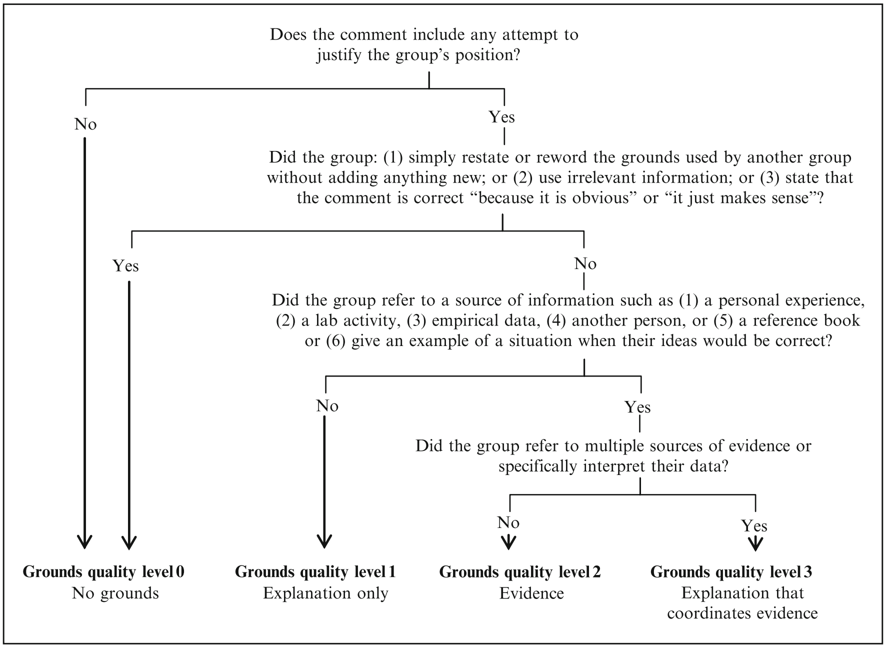

Fig. 10.1 Flow chart for coding grounds of a comment

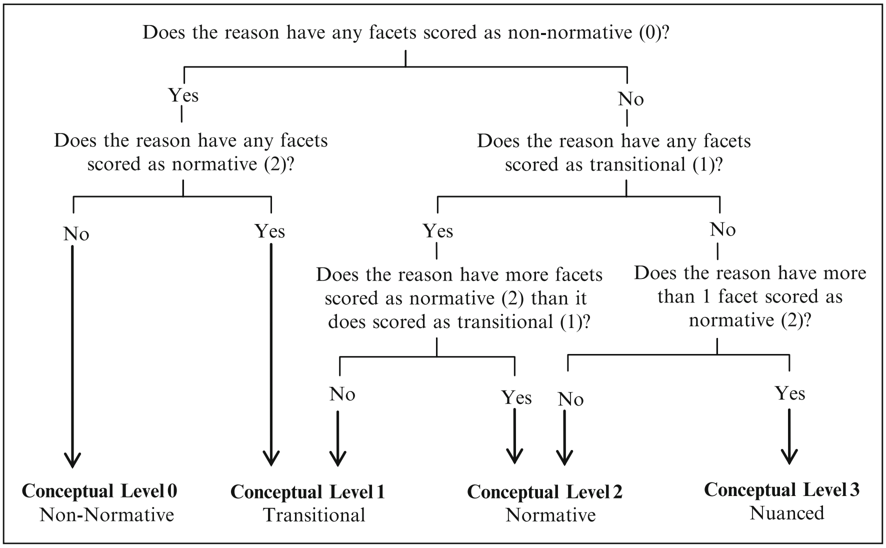

Fig. 10.2 Flow chart for coding the conceptual quality of a comment

Table 10.2 Example facets for coding conceptual quality of comment

Non-normative facets
Metal objects are above/below ambient temperatures by a great difference
Size/thickness affects final temperature
Good conductors/insulators keep heat on the surface *vs*. keeping heat/cold inside
Transitional facets
All objects in the same room will reach close temp (but not the same) as surroundings
Insulators "block, trap, or allow a small amount of" heat or cold (act like barriers)
Normative facets
Objects in the same room become the same temperature
If more heat flows into an object than out of it, its temp rises (or reverse)

Clark and Linn 2003). After coding the individual facets of a comment, the overall *conceptual quality* of a comment is determined through the series of binary decisions represented in the flow chart (Fig. 10.2). The flow chart assigns an overall *conceptual quality* score based on the frequency of non-normative, transitional, and normative facets found within the entire comment (see Table 10.2 for examples). As with the discussion of grounds above, coding of conceptual quality is not the focus of the current chapter, but full detail about this aspect of the coding scheme is available in Clark and Sampson (2008).

10.2.1.4 Coding the Level of Opposition within Discourse Episodes

After coding the individual comments, the framework then codes the larger episodes of discourse within which the comments occur. The framework considers an episode to be defined by each second-level comment (including its parent claim and its children). The framework characterizes the amount of conflict or *level of opposition* that takes place within an episode using the hierarchy outlined in Table 10.3. The framework defines high quality argumentation (*oppositional level 5*) as discourse that emphasizes the use of multiple rebuttals that challenge the interpretation of a phenomenon and the validity of the grounds used to support this interpretation. On the other hand, low quality argumentation is either non-oppositional (*oppositional level 0*) or consists of only claims and counter claims which do not attempt to challenge the validity of the other participants interpretation of the phenomenon (*oppositional level 1*). This scheme adapts the hierarchy outlined in Erduran et al. (2004) by incorporating the expanded definition of rebuttals outlined in Clark and Sampson (2007, 2008).

10.2.2 Using Sequential Analysis to Identify Discourse Patterns in Argumentation

Sequential analysis (Bakeman and Gottman 1997) has been used to analyze and model sequential links between behavioral events to determine how likely one given event is followed by another given event. Jeong (2004, 2005) developed the

Table 10.3 The overall quality of the argumentation and level of opposition that takes place within an episode is determined using a hierarchy based on opposition

Quality	Characteristics of the discourse
Level 5	Argumentation involving multiple rebuttals and at least one rebuttal that challenges the grounds used to support a claim
Level 4	Argumentation involving multiple rebuttals that challenge the thesis of a claim but does not include a rebuttal that challenges the grounds used to support a claim
Level 3	Argumentation involving claims or counter-claims with grounds but only a single rebuttal that challenges the thesis of a claim
Level 2	Argumentation involving claims or counter-claims with grounds but no rebuttals
Level 1	Argumentation involving a simple claim *versus* counter-claim with no grounds or rebuttals
Level 0	Non-oppositional

	+ARG	+BUT	+EXPL	+EVID	-ARG	-BUT	-EXPL	-EVID	Replies	No Replies	Givens	Reply Rate	% replies	% givens
+ARG	.02	.03	.25	.17	.00	.49	.00	.04	213	21	127	.83	.25	.10
+BUT	.00	.10	.05	.10	.00	.66	.06	.03	135	162	289	.44	.16	.23
+EXPL	.00	.02	.08	.15	.00	.67	.06	.02	52	64	112	.43	.06	.09
+EVID	.00	.00	.10	.13	.00	.71	.00	.06	31	50	84	.40	.04	.07
-ARG	.00	.48	.03	.02	.00	.02	.26	.19	174	21	124	.83	.20	.10
-BUT	.00	.61	.11	.02	.00	.08	.08	.09	157	185	328	.44	.18	.26
-EXPL	.00	.56	.13	.00	.00	.04	.17	.10	52	56	102	.45	.06	.08
-EVID	.00	.62	.05	.03	.00	.00	.15	.15	39	49	81	.40	.05	.06

Fig. 10.3 Transitional probability matrix produced by DAT

Discussion Analysis Tool (DAT) to compute the transitional probabilities between discourse moves observed in online debates. DAT has been used to produce transitional probability matrices to report, for example, the percentage of replies to stated arguments (ARG) that are challenges (BUT) *vs*. explanations (EXPL) *vs*. supporting evidence (EVID); and the percentage of replies to challenges that are counter-challenges *vs*. explanations *vs*. supporting evidence (see Fig. 10.3).

The matrix in Fig. 10.3 represents the message-response exchanges observed in an online debate. For example, the circled number indicates that 48% of all replies to the 124 opposing arguments (–ARG) were challenges (+BUT), for this group of students. The 124 opposing arguments (10% of all the discussion postings) elicited a total of 174 replies, approximately 20% of all the observed replies posted to the discussions. Only 21 of these 124 opposing arguments did not elicit any replies, and as a result, 83% of all opposing arguments elicited at least one or more replies.

DAT also produces a corresponding *z*-score matrix to identify and highlight transitional probabilities that are significantly higher/lower than expected

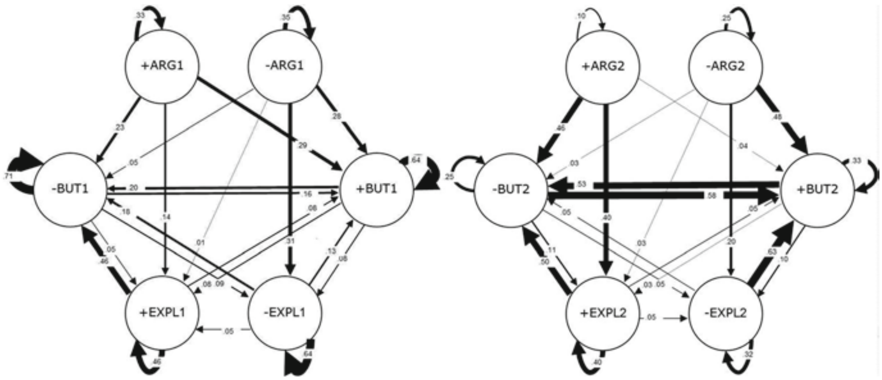

Fig. 10.4 Transitional state diagrams of response patterns produced by less intellectually open (*left diagram*) *vs*. more intellectually open students (*right diagram*)

probabilities. Probabilities that identify message-response sequences that can be considered to be behavioral "patterns" in an online debate. To visually and efficiently convey the complex data revealed in the transitional probability matrix, DAT converts the observed probabilities into transitional state diagrams (see Fig. 10.4). Potential differences in behavioral patterns between experimental groups—such as groups with students that have high *vs*. low in intellectual openness (Jeong 2007)—can be easily seen by juxtaposing state diagrams and observing the differences in the thickness of the links between events (signifying the strength of the transitional probabilities between given events). For example, a visual comparison of the two state diagrams in Fig. 10.4 shows that students that are more intellectually open (right diagram) exhibit a higher tendency to challenge one another's arguments (ARG→BUT) and counter-challenge one another's challenges (BUT→BUT) than students who are less intellectually open (left diagram).

To determine how an observed response pattern actually influences how often students post specific types of responses, DAT can be used to tabulate, for example, how many challenges are elicited by each argument, or how many explanations are elicited by each challenge. These scores can then be used to test for differences in the "mean response scores" – the mean number of challenges elicited per argument and the mean number of explanations elicited per challenge – between two or more experimental groups using statistical tests like the *t*-test and analysis of variance as demonstrated later in the case study.

10.3 A Sample Study and Analysis

To demonstrate the integration of the two methods described above, a case study was conducted to provide further insight into the findings of an earlier study (Clark et al. 2009). That study focused primarily on differences between two conditions in terms of differences in pre-post gains on the explanations that students constructed before and after the discussions. A brief analysis was also conducted, however, on the discourse moves within the discussions of each condition using the base coding methods outlined earlier from Clark and Sampson (2008). The findings from the analysis of the discourse moves was suggestive in that study, but not conclusive. By integrating the sequential analysis component, we hope to provide further insight into the findings of Clark, D'Angelo, and Menekse.

10.3.1 Data Sample

This analysis focuses on five ninth-grade integrated science classes taught by the same teacher at a public high school in a large metropolitan area in the southwestern United States. This was the participant group from our first trail in the original study. The teacher was an experienced teacher, but he had not worked with the online environment employed in this study or our research group prior to this study. The classes were typical ninth grade integrated science classes, labeled neither "honors" nor "remedial." Prior to this study, the students had conducted various inquiry projects, but had not explicitly studied dialogic argumentation within the curriculum of the class. The students worked on the project for approximately six class periods. The public school is located in a diverse city and has a roughly even distribution of boys and girls. The district is 58% Non-Hispanic White, 29% Hispanic, 6% Black, 6% Asian/Pacific Islander, and 1.4% American Indian/Alaska Native. The district categorizes 27% of the student population as economically disadvantaged. In total, there were 147 students, 38 discussion groups, and 2,160 discussion comments.

10.3.2 Instructional Context

The *personally-seeded discussion* system that is the focus of this case study is a customized asynchronous online discussion forum embedded within a Web-based Inquiry Science Environment (WISE) project called *Thermodyna mics: Probing Your Surrounding* (see http://wise.berkeley.edu). The *Thermodynamics: Probing your Surroundings* project consists of eight activities (see Fig. 10.5). In activities 1–5 students collected real time data about the temperatures of objects found inside the classroom and explore interactive simulations dealing with such ideas as heat transfer, thermal conductivity, and thermal sensation. As students worked through these activities they were prompted to record the data they gathered and describe the observations they made using the WISE note feature. We provide more detailed information about the project, the personally-seeded discussions, and the theoretical rationale for our approach in other publications (Clark 2004; Clark and Sampson 2007, 2008; Clark et al. 2009; Cuthbert et al. 2002).

In activity 6, students were asked to develop a principle that explained why objects that have been sitting in the same room for long periods of time often feel different.

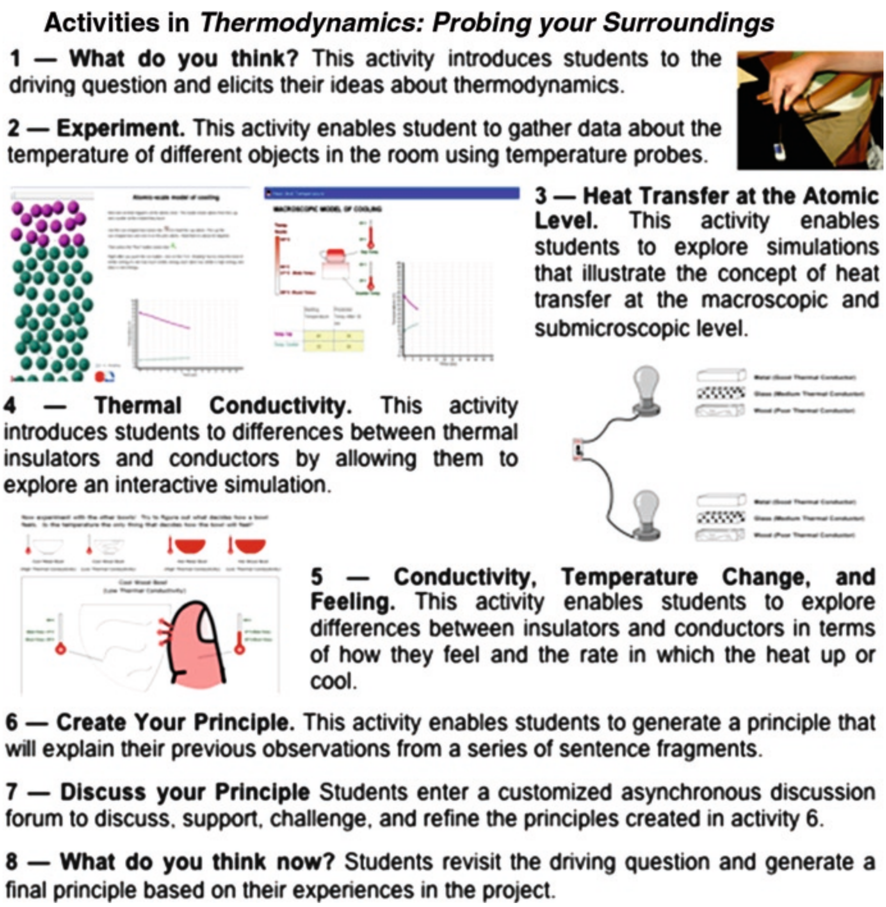

Fig. 10.5 Overview of the activities in the *Thermodynamics: Probing Your Surrounding* project

To scaffold students in this task and to ensure that students articulate their ideas clearly and focus on the salient issues of the problem, students use the *PrincipleMaker* interface. This interface allows students to use a pull-down menu format to create a principle from sentence fragments (see Fig. 10.6). The predefined phrases and elements include components of inaccurate principles that students typically use to describe heat, thermal equilibrium, and thermal conductivity that were identified through the misconceptions and conceptual change literature (e.g., Clough and Driver 1985; Erickson and Tiberghien 1985; Harrison et al. 1999) and an earlier thermodynamics curriculum development project (Clark 2000, 2004; Lewis 1996; Linn and Hsi 2000). This process serves multiple purposes. First, the pull-down format ensures that the students' conceptions of a phenomenon focus on the salient issues and are sufficiently elaborated to enable other students to note and discuss differences in their conceptions. Second, the pull-down menu format enables the discussion software to differentiate between students' principles so that

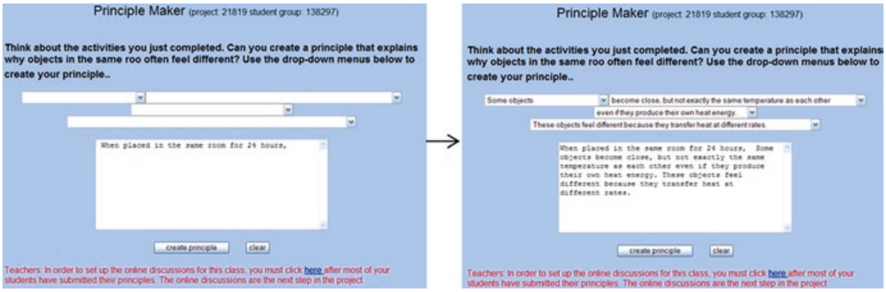

Fig. 10.6 In the PrincipleMaker explanation construction interface, students use a pull-down menu to construct an explanation from four sentence fragments that include common misconceptions

students can be automatically assigned to a discussion forum with other students who have constructed different principles to explain the same phenomenon.

Once students submit their principles, they move on to activity 7. In this activity, students participate in an asynchronous online discussion where they are encouraged to propose, support, critique, evaluate, and revise ideas. In order to foster argumentation, we designed the personally-seeded discussion software to set up and assign discussion forums to 3–5 students who have created different principles to explain the same phenomenon. This ensures that students are exposed to alternative interpretations of a given phenomenon.

10.3.3 Two Experimental Conditions

Each discussion group in the first trial of two in the original study (19 groups per condition) was randomly assigned to one of two conditions in terms of the nature of the seed comments in their discussion. The two conditions compared two seed-comment selection approaches for the discussion script. The personally-seeded groups received the explanations they constructed with the interface shown in Fig. 10.6 as their seed comments. Students in the augmented-preset groups received a pre-determined set of seed-comments constructed by the researchers using the same fragments supplied to the students in that interface. Table 10.4 shows two sets of seed comments to illustrate the difference between these two groups. The first set is the set received by the four-person augmented-preset groups. The second set is an example from a four-person group in the personally-seeded condition. The table also includes scoring information used in the original study to compare pre-post discussion gains in explanation quality.

In both conditions, the same initial scaffolding was used to enable students to explore the fragments that constituted the initial seed comments (the interface depicted in Fig. 10.6 prior to the discussions). Furthermore, the conflict schema approach was used in both conditions to for discussion groups that consisted of students with differing explanations. The two conditions diverge solely in terms of the

Table 10.4 Example sets of seed comments from discussion groups

Pre-selected Set of Seed Comments from the Augmented-Preset Groups (M = 7.50, SD = 3.70)	Score
When placed in the same room for 24 h, all objects become the same temperature as the room unless they produce their own heat energy. These objects feel different because they transfer heat at different rates.	12
When placed in the same room for 24 h, all objects become the same temperature as the room but only on their surface not inside them. These objects feel different because they transfer heat at different rates.	9
When placed in the same room for 24 h, objects that are good insulators stay at their original temperature regardless of the temperature of the room unless air can get inside them. These objects feel different because they are different temperature.	5
When placed in the same room for 24 h, some objects become close, but not exactly the same temperature as each other because they are made of different materials. These objects feel different because they are different temperature.	4
Example Set of Seed Comments from a Personally-Seeded Group (M = 8.00, SD = 2.16)	**Score**
When placed in the same room for 24 h, objects that are good conductors become the same temperature as the room unless they produce their own heat energy. These objects feel different because they are different temperature.	10
When placed in the same room for 24 h, hot objects become the same temperature as the room even if they produce their own heat energy. These objects feel different because they transfer heat at different rates.	9
When placed in the same room for 24 h, metal and glass objects become close, but not exactly the same temperature as the room unless they produce their own heat energy. These objects feel different because they transfer heat at different rates.	8
When placed in the same room for 24 h, all objects are at a different temperature than other objects in the same room because they are made of different materials. These objects feel different because they transfer heat at different rates.	5

third component (i.e., the nature of the initial seed comments). The augmented-preset seed comments were constructed to represent an optimized range of student misconceptions as opposed to including students' own explanations as the seed comments.

10.4 Discussion of Findings with Coding Scheme Only

Analysis of pre-post explanation gains in the original study showed that students in the augmented-preset condition demonstrated significant gains on their explanations from the first trial. A secondary analysis using the core coding scheme described earlier in this chapter and in Clark and Sampson (2008) was then conducted to provide additional insight into possible differences in the discussions in each condition that might have contributed to the observed differences in the pre-post gains in explanation quality. We now present an overview of the results from those analyses reported in the original study as a foundation for considering the potential value of using sequential analysis in tandem with the core coding scheme.

10.4.1 Conceptual Quality

An independent samples *t*–test showed that the mean conceptual quality level per episode of the comments in the augmented-preset condition ($M = 1.38$, $SD = 1.04$) was significantly higher than the mean in the personally-seeded condition ($M = 1.21$, $SD = 0.77$), $t(422) = 1.94$, $p < .05$. Clark, D'Angelo, and Menekse hypothesized that this might have resulted from the fact that students in the augmented-preset condition always received at least one fully normative explanation in a seed comment. Students in the personally-seeded condition received seed comments that were based solely on their own explanations – explanations that did not necessarily include fully normative explanations.

10.4.2 Grounds Quality and Frequency of Rebuttals

A few other noted differences in the discussion quality between the conditions suggested certain advantages of using the augmented-preset approach. These differences were not statistically significant, but followed trends from earlier studies and thus invited speculation. The mean grounds quality level of comments in the augmented-preset condition was higher, for example, than the mean in the personally-seeded condition. The students in the augmented-preset condition thus appeared to be more likely to include grounds for their statements as opposed to focusing on connecting the statements between individual participants. Similarly, the frequency of rebuttals in the augmented-preset condition was higher than the frequency of rebuttals in the personally-seeded condition. This may have been another function of the personal connections in the sense that students were less willing to rebut or contradict an explanation when it was "owned" by another person, in comparison to when the explanation was attributed to a non-present third party.

10.4.3 Discourse Moves

Figure 10.7 provides an overview of the numbers and types of discourse moves made by students in each condition. The overall patterns are very similar. One difference between the two groups is that the changing of claims occurred only once in the augmented-preset group (as opposed to 13 time in the personally-seeded group) – a meta-cognitive operation that would appear to be a critical part of the learning process. One possible explanation for this finding is that the students in the preset-augmented condition did not feel that they were examining their own ideas.

Lastly, students in the personally-seeded condition contributed higher word totals and numbers of comments (although not significantly higher in the current study) than the students in augmented-preset condition. Although this difference was not as large as in our previous studies, this finding suggests that students in the personally-seeded

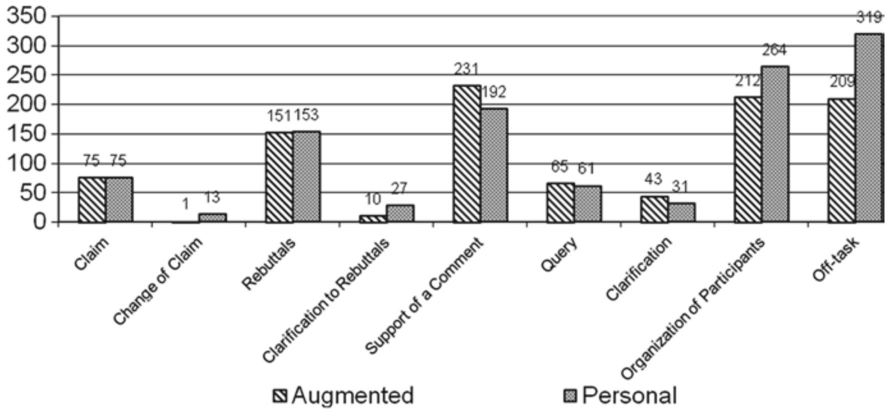

Fig. 10.7 Number and types of discourse moves in each condition

condition tend to type more than the students in the augmented-preset condition, which might suggest higher levels of engagement. As a result, the personally-seeded discussions appear to offer certain advantages as well as disadvantages when compared with the use of augmented-preset discussions along an "objectivity" *versus* "engagement" continuum.

10.4.4 Level of Opposition

Analysis of the structural level of opposition, however, showed no significant difference between the augment-preset and the personally-seeded discussion conditions in terms of the proportion of discourse episodes coded at each level of opposition, $\chi^2(5) = 2.83$, $p = .72$ (Fig. 10.8).

Another suggestive, but not statistically significant, difference between the augmented-preset and personally-seeded conditions involves the frequency of off-task comments. Approximately 28% of all comments in the personally-seeded condition were coded as off-task compared to 21% of the comments in the augmented-preset condition. This pattern was observed in a previous study (Clark et al. 2008). While both groups had many off-task comments about completely non-related topics, only the personally-seeded condition included comments that were personally focused in terms of applying social pressures to shift opinions (e.g., "we should pick mine" or "don't pick his").

One possible reason as to why the personally-seeded condition produced more off-task comments per student is because these students were more inclined or motivated to defend/support their own explanations and persuade others to accept their own explanations. While both groups produced off-task comments were not at all related to the topics (e.g., "nice haircut!"), only the personally-seeded condition produced comments (comment that were coded as off-task in this particular

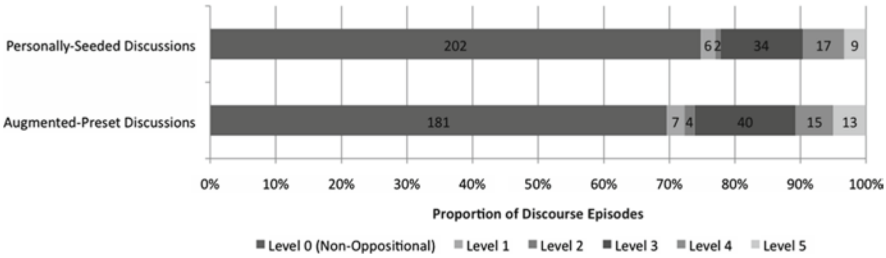

Fig. 10.8 Number of discourse episodes in the augmented-preset discussions and in the personally-seeded discussion at each level of opposition

study) that were aimed to persuade and draw group consensus. For example, students in the personally-seeded condition produced comments such as, "we should pick mine" or "don't pick his." These persuasive and consensus making types of comments may deserve a separate coding category in future studies.

Based on these analyses, the original study suggested that the personal embeddedness and engagement of the personally-seeded condition ultimately appeared to offer advantages as well as disadvantages compared to the augmented-preset condition along an "objectivity" *versus* "engagement" continuum. Overall, however, the core coding analysis of the ways students proposed, supported, evaluated, and revised ideas indicated that the augmented-preset condition seemed to be superior to the personally-seed discussions.

10.5 Discussion of Findings Using Sequential Analysis in Tandem with the Core Coding Scheme

10.5.1 Statistical Analysis

To sequentially analyze and identify differences in discourse patterns between conditions, the data (Fig. 10.9) used for this particular analysis consisted of 1,571 (2,160 total messages – 589 message were coded as "Other" when the categories were collapsed for the sequential analysis and thus omitted). Figures 10.9 and 10.11 present a breakdown of the observed frequencies and relative frequencies of messages (left column) and the most immediate and/or direct responses (at lag 0) to messages (top row). Messages that were posted in subsequent replies to an earlier message but separated by one or more previous responses (at lag 1 or more) were not examined in this study. The cell frequencies presented in bold identify response frequencies that were significantly higher than expected frequencies based on z-score tests at $p<.01$ (Fig. 10.10). The cell frequencies in italic and underlined identify response frequencies that were significantly lower than expected frequencies. The z-scores were computed for each possible event pairing while taking into account the differences in relative and observed frequencies of both given and target events. See Bakeman and

	CLAIMa	REBUTa	QUERYa	CLARIFYa	CHANGEa	SUPPORTa	CLAIMp	REBUTp	QUERYp	CLARIFYp	CHANGEp	SUPPORTp	Replies	No Replies	Givens	% replies	% givens
CLAIMa	1	**88**	16	0	0	**109**							214	1	75	.23	.05
REBUTa	0	**25**	14	**7**	0	**47**							93	58	151	.10	.10
QUERYa	0	7	**48**	0	**1**	15							71	29	108	.08	.07
CLARIFYa	0	2	0	**1**	0	1							4	5	10	.00	.01
CHANGEa	0	0	0	0	0	0							0	1	1	.00	.00
SUPPORTa	0	**22**	**17**	1	0	**50**							90	117	232	.10	.15
CLAIMp							1	**87**	12	*0*	4	***86***	190	0	75	.21	.05
REBUTp							0	19	**19**	**26**	**4**	***31***	99	47	153	.11	.10
QUERYp							0	4	**36**	0	2	*8*	50	32	92	.05	.06
CLARIFYp							0	6	3	1	0	**10**	20	10	27	.02	.02
CHANGEp							0	0	0	0	0	**2**	2	7	13	.00	.01
SUPPORTp							*0*	**22**	**14**	0	*0*	**47**	83	87	192	.09	.12
	1	144	95	9	1	222	1	138	84	27	10	184	916	837	1571		

Fig. 10.9 Frequency matrix from DAT with observed response frequencies to given messages. *Note*: a = message posted in augmented-preset condition; *p* = message posted in personal-seeded condition; bold values = higher than expected frequency; italic underlined values = lower than expected frequency

	CLAIMa	REBUTa	QUERYa	CLARIFYa	CHANGEa	SUPPORTa	CLAIMp	REBUTp	QUERYp	CLARIFYp	CHANGEp	SUPPORTp
CLAIMa	1.81	**11.66**	-1.59	-1.66	-0.55	**10.41**						
REBUTa	-0.34	**3.12**	1.56	**6.75**	-0.34	**6.24**						
QUERYa	-0.29	-1.41	**16.47**	-0.87	**3.45**	-0.64						
CLARIFYa	-0.07	1.89	-0.68	**4.88**	-0.07	0.04						
CHANGEa	0.00	0.00	0.00	0.00	0.00	-0.01						
SUPPORTa	-0.33	**2.39**	**2.79**	0.13	-0.33	**7.30**						
CLAIMp							1.96	**13.30**	-1.53	*-2.70*	1.51	**9.73**
REBUTp							-0.35	1.22	**3.66**	**14.52**	**2.99**	**2.95**
QUERYp							-0.24	-1.44	**15.83**	-1.27	2.04	-0.74
CLARIFYp							-0.15	1.89	0.91	0.55	-0.48	**3.38**
CHANGEp							-0.05	-0.60	-0.45	-0.25	-0.15	**2.82**
SUPPORTp							-0.32	**3.06**	**2.55**	-1.66	-1.00	**8.71**

Fig. 10.10 Z-score matrix revealing frequencies that were significantly higher (*bold values*) and lower (*italic, underlined*) than the expected frequency based on z-score tests at $p < .01$

Quera (1995, p.109) for more details on how the z-scores are computed in a way that takes into account the number of observed responses per category (marginal totals per column). As a result, the z-score values can be used as a means to operationally define what is to be considered (or not considered) a "discourse pattern".

Relative frequencies were computed from the frequency matrix with DAT and reported in a transitional probability matrix (Fig. 10.11). For example, the upper

	CLAIMa	REBUTa	QUERYa	CLARIFYa	CHANGEa	SUPPORTa	CLAIMp	REBUTp	QUERYp	CLARIFYp	CHANGEp	SUPPORTp
CLAIMa	.005	**.41**	.07	.00	.00	**.51**						
REBUTa	.00	**.27**	.15	**.08**	.00	**.51**						
QUERYa	.00	.10	**.68**	.00	**.01**	.21						
CLARIFYa	.00	.50	.00	**.25**	.00	.25						
CHANGEa	.00	.00	.00	.00	.00	.00						
SUPPORTa	.00	**.24**	**.19**	.01	.00	**.56**						
CLAIMp							.01	**.46**	.06	*.00*	.02	***.45***
REBUTp							.00	.19	**.19**	**.26**	**.04**	***.31***
QUERYp							.00	.08	**.72**	.00	.04	*.16*
CLARIFYp							.00	.30	.15	.05	.00	**.50**
CHANGEp							.00	.00	.00	.00	.00	***1.00***
SUPPORTp							*.00*	***.27***	**.17**	.00	*.00*	**.57**

Fig. 10.11 Transitional probability matrix revealing probabilities that were higher (*bold values*) and lower (*italic underlined values*) than expected probabilities

left corner of the transitional probability matrix shows that 41% of all responses to claims (CLa) in the augmented-preset group were rebuttals (RBa) in contrast to 46% in the *personally-seeded group* were rebuttals (CLp→RBp). To help reveal all the differences in response patterns between the two groups, DAT translated the relative frequencies into transitional state diagrams (Fig. 10.12). The top diagram reveals discourse patterns in the augmented-preset group, and the bottom diagram reveals discourse patterns (frequencies that were higher than the expected frequency) in the personally-seeded group.

10.5.2 Differences in Transitional Probabilities

A comparison of the transitional state diagrams in Fig. 10.12 reveals the response patterns between the two groups were quite similar overall. Nevertheless, the diagrams show that students using the augmented-preset threads were more likely to post responses to claims with supporting/grounding statements (51% of responses to claims) than students that used personally seeded threads (45%). More importantly, students using the augmented-present threads were also more likely to follow up and/or respond to rebuttals with supporting statements (51%) than students using personally-seeded threads (31%). These results suggest that when students use augmented-preset threads, they are more likely to support their ideas when they respond to claims or to rebuttals of a claim. A Chi-Square test showed that the distribution of support statements elicited across the six response categories were significantly different between the two groups $\chi^2(5)=14.1$, $p=.015$. These particular findings help to illuminate when and where students tend to support their ideas. The implications of this finding is that if students are encouraged and/or provided

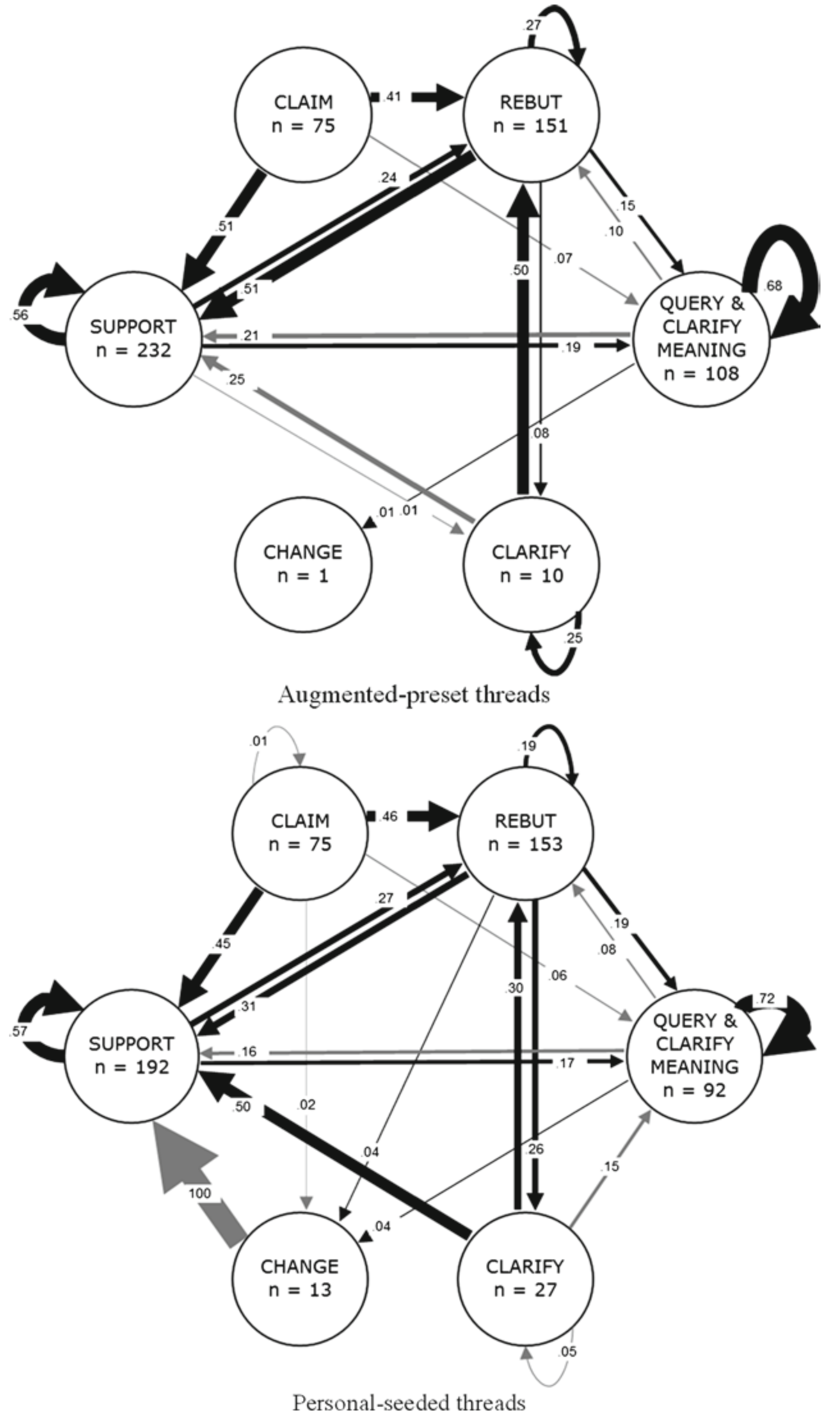

Fig. 10.12 Discourse patterns in augmented-preset *vs*. personally-seeded threads. Dark links = significantly higher than expected probabilities, Dotted links = significantly lower than expected probabilities

additional guidance on how to produce a greater number of rebuttals to each claim, we can expect to see the number of support statement increase as well.

10.5.3 Differences in the Mean Number of Responses Elicited Per Message

We conducted a 2 (conditions) × 4 (type of *oppositional* exchange) ANOVA to test for differences in the frequencies of four types of *oppositional* message-response exchanges – exchanges where rebuttals were posted in reply to claims, and where oppositional comments were posted in reply to rebuttals (i.e., claim → rebuttal, rebuttal → rebuttal, rebuttal → query meaning, rebuttal → change claim). We chose to select these four oppositional exchanges for this analysis based on the assumption that deeper inquiry is driven by the juxtaposition of differing viewpoints and to help reduce the chances of committing Type I error. We also conducted a 2 (conditions) × 6 (*supportive comments* posted in reply to claims, rebuttals, queries, clarifications, change claims, and supportive comments) ANOVA to test for differences across the six primary types of supportive exchanges (i.e., claim → supporting comment, rebut → support, query → support, clarify rebuttal → support, change claim → support, support → support) based on the differences noted in the transitional state diagrams.

10.5.3.1 Oppositional Exchanges

No significant differences were found between conditions in the number of responses posted across the four types of oppositional exchanges, $F(1, 1{,}054) = .00$, $p = .982$. The results of the sequential analysis revealed no indications that one condition lead to higher levels of argumentation in terms of the oppositional exchanges than the other condition. Significant differences were found in the number of responses elicited per message between the four different types of exchanges, however, independent of condition, $F(3, 1{,}054) = 211.26$, $p = .000$. In other words, certain types of exchanges tended to elicit more responses than other types of exchanges. Claims elicited on average 1.17 rebuttals ($STD = 1.06$, $n = 150$), rebuttals elicited .14 counter rebuttals ($STD = .396$, $n = 304$),.11 queries ($STD = .332$, $n = 304$), and .01 change claims ($STD = .114$, $n = 304$). No interaction was found between oppositional exchange type and condition, $F(3, 1{,}054) = .36$, $p = .780$.

10.5.3.2 Supportive Exchanges

No significant differences were found between conditions in the number of responses posted across the six supportive exchanges examined in this study, $F(1, 1{,}117) = .00$, $p = .983$. We found no indication that one condition lead to higher levels of supportive exchanges (as opposed to argumentative exchanges) than the other condition. Significant differences were found in the number of responses elicited per message between the six supportive exchanges independent of condition, $F(5, 1{,}117)=90.86$, $p = .000$. The average number of supporting comments posted in reply to claims was 1.30 ($STD=1.11$, $n=75$ claims), .26 for rebuttals ($STD = .52$, $n=304$), .12 for queries ($STD = .22$, $n=200$), .30 for clarify rebuttals ($STD = .46$, $n=37$), .14 for change claims ($STD = .36$, $n=14$), and .23 for supporting comment ($STD = .47$, $n=424$).

We did find an interaction between type of supportive exchanges and conditions (Table 10.5), $F(5, 1{,}117)=2.42$, $p = .034$. In the augmented-preset group, students posted 27% more supportive comments in reply to claims, 54% more in reply to rebuttals, and 60% to queries. In contrast, students in the personally-seeded group posted 270% more supportive comments in reply to clarifying rebuttals, and 14% more in reply to supportive comments. We conducted additional analysis and found that: (a) claims, rebuttals, and queries (messages that elicited more supporting comments in the augmented-preset group than the personally-seeded group) were posted on average at 3.00 ($STD=2.07$, $n=654$) thread levels deep in discussion threads; (b) clarify rebuttals, change claims, and supportive comments (messages that elicited more supporting comments in the personally-seeded group) were posted on average at 3.23 ($STD=1.75$, $n=475$) levels deep; and (c) that this observed difference in thread level was statistically

Table 10.5 Mean number of supportive comments posted in reply to discourse moves between groups

Exchange	Group	Mean	std	n
Claim-support	Preset	1.45	1.26	75
	Personal	1.15	.93	75
	Total	1.30	1.11	150
Rebut-support	Preset	.31	.57	151
	Personal	.20	.46	153
	Total	.26	.52	304
Query-support	Preset	.14	.35	108
	Personal	.09	.28	92
	Total	.12	.32	200
ClarifyRebuttal-support	Preset	.10	.32	10
	Personal	.37	.49	27
	Total	.30	.46	37
Change-support	Preset	.00		1
	Personal	.15	.38	13
	Total	.14	.36	14
Support-support	Preset	.22	.46	232
	Personal	.24	.48	192
	Total	.23	.47	424
Total	Preset	.38	.76	577
	Personal	.33	.62	552
	Total	.36	.69	1,129

significant, $t(1,127) = -1.96$, $p = .05$. This finding suggests that students in the augmented-preset threads tended to reply with supportive comments to comment types that occurred earlier in a discussion thread, where as students in the personally-seeded discussions tended to reply with supportive comments to comment types that occurred later in a discussion thread.

10.6 Affordances of Using Both Methods in Tandem

In summary, the analysis performed with the core coding scheme alone suggested that (a) augmented-preset threads produced comments with higher conceptual quality (or more normative explanations as defined in Clark and Sampson 2008); (b) augmented-preset threads *may* have helped to produce more grounded claims; (c) augmented-preset threads *may* have helped to produce more rebuttals on the grounds of each claim; and (d) no differences were found in the proportion of episodes across each of the five levels of opposition (Table 10.6). The sequential analysis then revealed that (a) the patterns of discourse between the groups were overall very similar in structure; (b) there were no significant differences in the number of oppositional exchanges produced by students between the groups; (c) the number of responses posted in reply to each message depended heavily on the function or type of message; and (d) the time and place where students reply with supporting comments depends both on the type of message they are replying to and whether students are using augmented-preset *versus* personally-seeded discussions. Students using augmented-preset threads, in other words, were more likely than students using the personally-seeded threads to respond to claims and rebuttals with supporting statements. In this case, using both methods in tandem enabled us to: (a) pinpoint where, when, why, and/or how particular types of discourse moves of interest are elicited within the course of a conversation; and (b) identify where and how changes in the discourse process can be made to help increase the frequency of discourse moves of particular interest. In all, the sequential analysis revealed patterns that were more or less consistent with the previous findings reported by Clark and Sampson and also provided a quantitative and process-oriented approach to describing the nature and quality of argumentation in these different discussion forums (see Table 10.6).

Overall, our sequential analysis of the data sequential analysis produced potential explanations for the earlier findings identified with the core coding scheme and provided further insights into the patterns that emerged within the students' discourse. We therefore believe that using the core coding scheme in tandem with sequential analysis can provide useful insights into online discourse by providing visual representations (including quantitative measures) of the discourse process in ways that can help use better understand of how online discussion forums (both asynchronous and synchronous) affect the way discourse unfolds over time (when examined at the micro level) and how changes in processes help to produce quality argumentation.

Table 10.6 Findings on the effects of using augmented-preset discussion threads using core coding scheme and sequential analysis

Outcome	Core coding scheme	Sequential analysis
Conceptual quality	Higher	*Not investigated*
Level of grounding	Higher (but NS)	*Not investigated*
Rebuttal frequencies	More rebuttals (but NS)	State diagrams revealed no major differences in patterns (transitional probabilities with significant z-scores) in the messages that preceded and followed each rebuttal.
Level of opposition	No difference in proportion of episodes across each level of opposition (based on holistic evaluation of entire discussion thread)	Oppositional exchanges – No difference in mean number of oppositional responses to claims and rebuttals.
		Common patterns – Most response patterns identified from the sequential analysis were shared between groups. These patterns included the following exchanges:
		CLAIM – REBUT
		CLAIM – SUPPORT
		REBUT – REBUT
		REBUT – QUERY
		REBUT – SUPPORT
		QUERY – QUERY
		SUPPORTING – SUPPORT
		Unique patterns – Analysis of response patterns reveal that students in the augmented-preset group tended to respond back with further clarifications of rebuttals, where as in the personally-seeded group tended to respond to the clarification of rebuttals with supporting comments,
		Supportive exchanges – Differences in the way students posted supportive comments between the two groups depended on the function or discourse move that triggered students' responses.

10.7 Directions for Future Research

In future studies, we intend to examine (a) how conceptual quality correlates with specific discourse patterns, and (b) how and to what extent specific patterns help promote and/or explain observed differences in conceptual quality. We plan to explore, for example, the extent to which high *versus* low levels of oppositional exchanges trigger/elicit subsequent comments that are more normative and/or nuanced. Given that there were few differences in response patterns observed between the two groups, a more detailed analysis of the response patterns produced by student groups *within* the augmented-preset condition might help shed light on discourse patterns that promote conceptual quality. For example, a Markov analysis can be applied to our data to determine if there are significant differences in the frequency of particular three-event chains of discourse moves (as opposed to two-event chains) that distinguish one group from the other – Markov chains that might help to explain observed differences in the quality of the group performance overall.

Our future work will also explore the relationship between response patterns and the level of grounds students include in their arguments. To examine this relationship, the Clark and Sampson coding scheme will be expanded to differentiate responses that clarify, request clarification, and support in terms of whether they focus on the grounds or the thesis of the parent comment. At present, the Clark and Sampson coding scheme only differentiates rebuttals in terms of whether or not they focus on the thesis or the grounds of the parent comment. With an elaborated coding scheme that makes this differentiation for other comment types, we will apply sequential analysis to determine the percentage of responses to comments that focus on the thesis *versus* the grounds of a claim. Next, we can determine to what extent the ratio of focus on thesis *versus* grounds affects the level of grounds observed across all messages posted within a discussion thread (and perhaps across both experimental groups). Given the assumption that students are working under limited time and resources, one can test the claim that these two goals of focusing on the thesis and the grounds of comments are working in competition or synergy with one another. Any observed tendencies in students' responses that pursue one particular goal may have an adverse effect on the extent to which they are able to accomplish other goals. The observed response tendencies can then be compared between conditions to explain any observed differences in grounding.

One potential constraint with sequential analysis (when used to examine adjacent message and responses to messages), however, is that each observed response must be explicitly mapped or threaded to the correct message stated previously within a conversational thread. Students often post responses that perform multiple discourse moves that address multiple comments from multiple messages (i.e., messages posted immediately prior to the response and posted earlier in the message thread). One way to address this limitation is to modify both the coding scheme and coding procedures. For example, we can: (a) expand the coding scheme by assigning one code for 'a rebuttal against the thesis of a *claim*' and another code for 'a rebuttal against the thesis of a *rebuttal*'; and (b) parse messages that perform multiple discourse moves into separate units and assign individual codes to each unit. Another alternative is to integrate pre-specified prompts into the discussion board to constrain each posting to respond only to the parent message while using one and only one discourse move. Although each of these solutions presents its own set of limitations or issues, these types of changes can potentially increase the accuracy and precision of the state diagrams resulting from a sequential analysis of the students' conversations. This increased accuracy and precision would support a more detailed examination of the relationships between discourse processes, conceptual quality, and level of grounding.

One limitation of the DAT software is that the number of discourse moves presented in each state diagram is limited to a maximum of six discourse moves. The software tool will require further changes so that it can generate state diagrams which can convey transitional probabilities between larger numbers of discourse moves to conduct some of the future studies described above. Furthermore, the transitional diagrams generated with DAT will need to also convey the probabilities in which each message elicits no response. In doing so, the observed transitional probabilities (response patterns) might provide more accurate explanations for the observed differences in mean response scores (e.g. the average number of challenges posted in reply to a claim). The software will also need to include a mechanism that enables the viewer to: (a) flip and superimpose one state diagram over another diagram to make it easier to visualize and identify the similarities and differences between diagrams (particularly with diagrams containing large numbers of discourse moves); and (b) aggregate the diagrams into one diagram to reveal the similarities and differences (using links with varied colors and/or gray scale) with respect to or relative to one selected diagram. Tools for aggregating data across matrices and superimposing transitional state diagrams over another diagram can be found in the software application called jMAP (Jeong 2008). Tools like this could be integrated into DAT to facilitate the comparison of larger and more complex state diagrams.

We would also make the following additional recommendations for future research: (a) expand the analysis to measure the frequency of three-event sequences to determine whether some event pairs are more effective in eliciting desired responses than other event pairs; (b) analyze the discourse between experts/teachers and identify sequences that distinguish experts from novices using multidimensional scaling; (c) test and validate process models across variants of the task using new message codes and labels to facilitate discussions and to identify new patterns of interaction that support group performance; (d) examine how specific scaffolds and instructional strategies affect the way discourse patterns change over time (learning trajectories) by visually flipping and superimposing state diagrams of discourse patterns observed across different time periods over a target state diagrams depicting discourse patterns exhibited in the discourse between experts and teachers; and (e) assess scientific explanations by examining students' causal loop diagrams and use tools like jMAP and DAT to examine how discourse patterns trigger changes in students' causal diagrams/understanding that converge toward expert diagrams/understanding.

This chapter, overall, demonstrates the application of Clark and Sampson's coding scheme for capturing the processes of scientific argumentation and how sequential analysis can be used in tandem as a way to provide both quantitative and qualitative descriptions of discourse processes in instructional contexts. We hope that the ideas presented here will form the basis of a new process-oriented framework for measuring discourse and argumentation in CSCL environments and developing new process-oriented methods to support, monitor, evaluate, and improve student learning and performance.

References

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

Bakeman, R., & Quera, V. (1995). *Analyzing interaction: Sequential analysis with SDIS and GSEQ* (p. 109). Cambridge University Press: Cambridge.

Clark, D. B. (2000). Scaffolding knowledge integration through curricular depth. Unpublished doctoral dissertation, University of California, Berkeley.

Clark, D. B. (2006). Longitudinal conceptual change in students' understanding of thermal equilibrium: An examination of the process of conceptual restructuring. *Cognition and Instruction, 24*(4), 467–563.

Clark, D. B., D'Angelo, C. M., & Menekse, M. (2009). Initial structuring of online discussions to improve learning and argumentation: Incorporating students' own explanations as seed comments versus an augmented-preset approach to seeding discussions. *Journal of Science Education and Technology, 18*(4), 321–333.

Clark, D. B., & Linn, M. C. (2003). Scaffolding knowledge integration through curricular depth. *Journal of the Learning Sciences, 12*(4), 451–494.

Clark, D. B. (2004). Hands-on investigation in Internet environments: Teaching thermal equilibrium. In M. C. Linn, E. A. Davis, & P. Bell (Eds.), *Internet environments for science education* (pp. 175–200). Mahwah, NJ: Lawrence Erlbaum Associates.

Clark, D. B., & Sampson, V. (2005). Analyzing the quality of argumentation supported by personally-seeded discussions. Paper presented at the computer supported collaborative learning conference, Taipei, Taiwan.

Clark, D. B., & Sampson, V. D. (2007). Personally-seeded discussions to scaffold online argumentation. *International Journal of Science Education, 29*(3), 253–277.

Clark, D., Sampson, V., Weinberger, A., & Erkens, G. (2007). Analytic frameworks for assessing dialogic argumentation in online learning environments. *Educational Psychology Review, 19*(3), 343–374.

Clark, D. B., & Sampson, V. (2008). Assessing dialogic argumentation in online environments to relate structure, grounds, and conceptual quality. *Journal of Research in Science Teaching, 45*(3), 293–321.

Clark, D. B., Schleigh, S. P., Menekse, M., D'Angelo, C. M., & Sampson, S. (2008). Improving the quality of student argumentation through the initial structuring of online discussions. Paper presented as part of symposium organized by William Sandoval for the American Educational Research Association (AERA) 2008 meeting. New York.

Clough, E. E., & Driver, R. (1985). Secondary students' conceptions of the conduction of heat: Bringing together scientific and personal views. *Physical Educator, 20*, 176–182.

Cuthbert, A. J., Clark, D. B., & Linn, M. C. (2002). WISE learning communities: Design considerations. In K. A. Renninger & W. Shumar (Eds.), *Building virtual communities: Learning and change in cyberspace* (pp. 215–246). Cambridge: Cambridge University Press.

deVries, E., Lund, K., & Baker, M. (2002). Computer-mediated epistemic dialogue: Explanation and argumentation as vehicles for understanding scientific notions. *Journal of the Learning Sciences, 11*(1), 63–103.

Erduran, S., Simon, S., & Osborne, J. (2004). TAPping into argumentation: Developments in the application of Toulmin's argument pattern for studying science discourse. *Science & Education, 88*, 915–933.

Erickson, G., & Tiberghien, A. (1985). Heat and temperature. In R. Driver, E. Guesne, & A. Tiberghien (Eds.), *Children's ideas in science* (pp. 52–83). Philadelphia: Open University Press.

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

Harrison, A. G., Grayson, D. J., & Treagust, D. F. (1999). Investigating a grade 11 student's evolving conceptions of heat and temperature. *Journal of Research in Science Teaching, 36*(1), 55–87.

Janssen, J., Erkens, G., Jaspers, J., & Kanselaar, G. (2007). Visualization of participation: Does it contribute to successful computer-supported collaborative learning? *Computers & Education, 49*, 1037–1065.

Jeong, A. (2004). Discussion analysis tools. Retrieved March 10 2009. http://mailer.fsu.edu/~ajeong/DAT

Jeong, A. (2005). A guide to analyzing message-response sequences and group interaction patterns in computer-mediated communication. *Distance Education, 26*(3), 367–383. Retrieved December 2008 http://garnet.fsu.edu/~ajeong/research/Jeong2005_AnalyzingInteraction.pdf.

Jeong, A. (2007). The effects of intellectual openness and gender on critical thinking processes in computer-supported collaborative argumentation. *Journal of Distance Education, 22*(1), 1–18.

Jeong, A. (2008). jMAP. Retrieved March 10 2009. http://jmap.wikispaces.com

Jermann, P., & Dillenbourg, P. (2003). Elaborating new arguments through a CSCL script. In J. Andriessen, M. Baker, & D. Suthers (Eds.), *Arguing to learn: Confronting cognitions in computer-supported collaborative learning environments* (pp. 205–226). Dordrecht: Kluwer Academic.

Lewis, E. L. (1996). Conceptual change among middle school students studying elementary thermodynamics. *Journal of Science Education and Technology, 5*(1), 3–31.

Linn, M. C., & Hsi, S. (2000). *Computers, teachers, peers: Science learning partners*. Mahwah: Lawrence Erlbaum.

Rourke, L., Anderson, T., Garrison, D. R., & Archer, W. (2001). Methodological issues in the content analysis of computer conference transcripts. *International Journal of Artificial Intelligence in Education, 12*, 8–22. Retrieved July 09, 2005, from http://aied.inf.ed.ac.uk/members01/archive/vol_12/rourke/full.html.

Schwarz, B. B. & Glassner, A. (in press). The role of CSCL argumentative environments for broadening and deepening understanding of the space of debate. In R. Saljo (Ed.), Information technologies and transformation of knowledge.

Chapter 11
Is the Whole Greater than the Sum of Its Parts? Explaining the Role of Individual Learning and Group Processes in CSCL

Susanne P. Lajoie

Computer-supported collaborative learning research considers learning as it occurs in complex situations, where human thought and action occurs in response to the environment (the task and social group) with respect to how complex contexts provide opportunities for integrating information from multiple sources (Clancey 1997; Greeno 1998; von Glaserfeld 1995). Accounting for changes or progressions in learning in such complex situations is difficult to tease apart due to the interdependency of the task, the group and the individual. The contributions in part two of this volume speak directly to this complexity by providing the readers with valuable examples of mixed method approaches to understanding the role of the individual learning as well as the group processing that occurs in CSCL research. Two chapters speak directly to the use of multilevel analyses (hierarchical linear modeling-HLM) as a way to accurately isolate statistically the contributions of the individual as well the group in CSCL research (Janssen, Erkens, Kanselaar, & Kirschner; Stylianou-Georgiou, Papanastasiou & Puntambekar). The other chapters speak directly to the temporal and sequential nature of knowledge development or argumentation, and describe data mining techniques and sequential pattern recognition as a step in describing changes in the group performance over time (Reimann, Yacef, & Kay; Jeong, Clark, Sampson, & Menekse). Jeong et al. discuss ways in which specific discourse acts may predict changes in argumentation. In the sections below I refer to some of the highlights of these chapters.

11.1 The Power of Multilevel Analysis (MLA) in CSCL Research: N Is the Answer

Janssen, Erkens, Kanselaar, and Kirschner provide CSCL researchers with an in depth description of the statistical concerns with the data analysis techniques being used to identify contributions of the individual and the group to the CSCL

S.P. Lajoie (✉)
Department of Educational and Counselling Psychology, McGill University
e-mail: susanne.lajoie@mcgill.ca

S. Puntambekar et al. (eds.), *Analyzing Interactions in CSCL,* Computer-Supported Collaborative Learning Series, DOI 10.1007/978-1-4419-7710-6_11,

experience. They review three major concerns. The first being hierarchical nesting, the second non-independence of dependent variables and the third being the differing units of analysis. Hierarchical nesting refers to the fact that individuals are nested within groups and groups within classrooms and thus they are not independent of one another. This is a problem when looking at differences between individuals or groups or classes on dependent variables since the individual is not independent of group, and the group is not independent from the classroom influence. MLA can handle the different nesting levels.

The non-independence of dependent variables issue refers to the fact that students in a group may be more similar to each other than to those students in another group (Kenny et al. 2002). Individuals in the group can have either a negative or positive effect on each other but they definitely have an influence on the learning situation and thus there is the non-independence issue. Janssen et al. suggest using the intraclass correlation coefficient to determine the amount of variance that is accounted for by the group and if it is significant and large enough one might assume there is bias and that either type 1 or 2 errors may result in statistical tests. The other issue that Janssen et al. discuss is the trouble with mixing different units of analysis, mixing of individual and group as independent variables and only measuring on dependent variables pertaining to one or the other.

Janssen et al. propose MLA as an answer to such problems and provide concrete examples comparing analyses results from traditional statistical methods with MLA. More specifically they demonstrate the typical ways people deal with such issues, ignoring the non-independence issue, aggregating or disaggregating individual data into or from group data. Their examples are clear and coherent demonstrations of how people have tried to deal with statistical concerns and the resultant outcomes of such analyses. Interestingly enough they demonstrate three ways to manage the same data with conflicting results demonstrating the prevalence of type 1 or 2 errors in methods that are not using MLA. MLA is a superior method for addressing the dependency issues on the individual, group and classroom levels (Strijbos and Fischer 2007).

Janssen et al. did an impressive presentation of what CSCL researchers should consider when selecting their analysis techniques. Regrettably, the authors point out that MLA is not the Holy Grail for all CSCL research. The greatest hurdle is that the recommended sample size for MLA is a minimum of 50 groups. A great deal of CSCL research is done with much smaller sample sizes that utilize mixed methods approaches that combine quantitative and qualitative analyses to present “rich descriptions,” of what is occurring in such contexts. Qualitative analyses of 50 groups could be quite cumbersome. Furthermore, many CSCL researchers are interested in documenting changes in patterns of learning within the group context over time, as we saw with the Reimann et al. and the Jeong et al. papers. Rich qualitative analyses can lead to effective data mining techniques and even sequential analyses techniques. However, if the question of whether or not to use MLA depends on sample size then N becomes the answer. CSCL researchers are restricted in the choices they can make based on their questions. Their respective methodology and analyses must fit their questions.

11.2 The Sum Is Better than Its Parts

The chapter by Stylianou-Georgiou et al. provide a very rich example of how MLA can isolate the contributions of the individual and the group in a CSCL situation to specific dependent measures. They provide an excellent example of how to examine the unit of analysis of both the individual and the group in the context of a collaborative design project in science, using CoMPASS. CoMPASS is a science hypertext system, where grade 6 students use this system as a resource tool to study science ideas pertinent to a design project on pulleys. Mixed ability groups were assigned to either a supported condition (where metanavigational prompts were provided dynamically based on navigation patterns logged by the computer) or an unsupported condition. Computer trace files of navigational strategies were used to determine the types of prompts to give groups to improve their understanding in the supported condition. Comparisons were made of performance and understanding of those in the supported versus non-supported group.

Using a multilevel HLM analysis Stylianou-Georgiou et al. were able to examine both the individual and the group in their experiment. They examine the cognitive attributes of the individual in two ways. First, they explore individual differences in metacognitive awareness using the Metacognitive Awareness of Reading Strategies Inventory (MARSI, Mokhtari and Reichard 2002). Second, they looked at how much each individual learned as a function of the CSCL by administering a post assessment of scientific understanding that required the creation of concept maps along with explanations and connections among concepts. Oddly enough, those individuals with low levels of metacognitive ability as identified by the MARSI, did not benefit from the metanavigational supports, as evidenced by their depth of understanding scores. However, using groups as a unit of analysis, they found that supported groups performed better than unsupported groups. Groups were examined by looking at their goal setting and navigation patterns taken to accomplish their goal. Group goals for navigation were examined along with the number of coherent transitions between concepts investigated to reach their goal. The groups in the supported condition outperformed the non-supported groups. In this case the whole is better than the sum of its parts since the supported groups did better than the supported individuals in the groups.

The analytical techniques used in this research were effective for isolating distinct contributions of the individual and the group. Stylianou-Georgiou et al. state that they will consider different cognitive attributes in the future, such as prior knowledge as a predictor of understanding, rather than or in addition to metacognitive ability. Given that their outcome measure was depth of understanding rather than changes in metacognitive ability an assessment of prior knowledge would be most appropriate. However, if the researchers are particularly interested in changes in metacognitive activity there are still alternatives that can be explored. For instance, it might be interesting to look at the results of the MARSI to determine the right level of metanavigational prompts that could be adapted to the needs of the individuals who constituted the group in question. Currently, the prompts are designed to adapt to the group's navigational paths. These prompts are obviously necessary and significant since there was a condition effect showing that groups in the supported condition did better than those in the unsupported conditions. However, the prompts could be made more meaningful to the individuals in the group if designed based on the MARSI results, potentially resulting in both a stronger group effect and individual gains on the outcome measures in question.

11.3 Mirrors for Metacognition: Visual Aids to See Oneself and Others in the Context of Problem Solving

Visual representations of the individual and group activity are powerful ways to disambiguate the role of individuals in a group, or the roles of different groups in contributing to the overall classroom experience. Chapters by Reimann et al. and Jeong et al. both use visual representations of student learning in novel ways. Visual representations can be used in two ways: by researchers to determine what is going on in a CSCL experience, or by learners to reflect on their own actions in comparison to others. When used with learners they can be called mirroring devices, that mirror an individual's actions in the context of a group learning situation (Kay et al. 2006). These mirrors can be useful devices for learners to reflect on their own performance, to become more aware of what they know or do not understand in the context of the group activity. The chapter by Reimann et al. provides insights into the use of data mining techniques in CSCL situations paired with graphical interface tools that model the learner and group activity. Alternatively, Jeong et al. provide rich examples of how transitional state diagrams can be used to document group differences in argumentation patterns. These representations are useful to researchers for making sense of where the group differences lie. It might be interesting to use the transitional state diagrams with students as well, in the same way that Kay and her colleagues use representations as mirroring and reflection tools. It could be advantageous for individuals to see how active they are in the argumentation activity and how the type of discourse evolves over time and how they themselves influence the nature of the activity. Reflecting on these representations can help individuals see when they are silent as well as when they are intrusive. In both such situations individuals may not have an influence on the groups activity and may lose control of their own learning through argumentation.

Reimann et al. provide a thoughtful review of the design of process mining techniques that can reveal "change processes" that are prevalent in the group. They discuss the need for theory-driven approaches to understanding collaboration and they consider different theoretical paradigms. Given their interest in documenting changes in learner activity over time they look to developmental theories since they deal with change and maturation of skills. However, they conclude that developmental theories too often discuss changes in stages that occur in a fixed order and that in group learning situations it is difficult to determine the contributions of each individual to the resultant shift in group-learning over time. Activity systems theory often looks at changes in the phases of group processing as a function of the task requirements but they do not conclude that changes occur in any fixed order. They concur that Engestrom's (1987) activity system theory is most appropriate for CSCL designs, where the group works together in a coordinated manner to reach a common goal, through planning and the appropriate division of labour. They argue that data mining may reveal patterns across levels of performance and that it would be unlikely that the emergence of group learning occurs in specific sequences that occur in a fixed order.

Reimann et al. argue for the need for theory-driven approaches to data mining in CSCL; however, they demonstrate the difficulties involved given the complexities of documenting learning through collective actions. They argue that one method of group learning is that the group learns through deliberate practice (Ericsson et al. 1993) by critiquing and reflecting on their interactions and their collective reactions to feedback from a tutor. This fits well with Clark and Sampson's (2007, 2008) discussion of coding argumentation based on levels of opposition, whereby good arguments are developed when groups are open enough to rebut each other's propositions and propose new ones. The group itself may be greater than the sum of its parts as seen in the Stylianou-Georgiou chapter. Reimann et al. discuss the many reasons that groups may differ based on the lifespan of the group. For example, the group itself has a history, and it may learn from its own experiences. The group also has a future, depending on the purpose and consequences of the group interaction different levels of learning and engagement may occur. The group's lifespan, be it short, medium or long term, may influence the ways in which the group works together. Reimann et al. describe the temporal nature of group learning, as it unfolds over time, is dependent on the groups previous actions, it is cumulative in that current knowledge influences future knowledge and it has an anticipated future based on interpreted past experiences. It is this temporal nature that must be considered when developing methods for analyzing group process data.

The concept of change and temporality can be translated into methods for analyzing event sequences and developing mechanisms to automatically identify sequence patterns and models with data mining techniques. Reimann et al. provide examples of both top-down theory-driven data mining of group interactions and bottom-up approaches that are non-theory-driven and used to detect patterns in the data. The top down approach used Salas et al. (2005) theory that identified five elements of successful group work (leadership, mutual performance monitoring, backup behavior, adaptability and team orientation), to guide the interpretation of the patterns found in an online course in computer science (Kay et al. 2006). Using this approach Kay et al. were able to isolate successful and unsuccessful group performance by defining temporal patterns that reflected different levels of group learning based on Salas et al.'s theory of group learning. After identifying such patterns they used them in their automatic data mining techniques. This is solid theory driven research that makes the interpretation of the data mining algorithms easier to use and more intuitive since it fits with the chosen theory. However, Reimann et al. point out that it is challenging to relate the theoretical constructs of pedagogical and psychological theories to data mining techniques since they were not developed with that goal in mind. It is even more difficult to automate collaboration pattern recognition since it is not clear what constitutes good or poor group performance. Consequently, bottom-up discovery approaches are useful methods to determine group interaction patterns. One way to explore group differences is to examine the temporal sequence of events and relate them to outcome data (i.e. grades) to see which sequences correspond to better performance on a complementary task. The different patterns that emerge can be used diagnostically to tell why some groups perform well and some do not.

Reimann et al. caution that data mining has considerable potential for constructing process models but that the quality of the model depends on the quality and representativeness of the data in which it is constructed and caution against the garbage-in garbage-out problem. They basically demonstrate that there is still a human element needed to this type of data mining, where human raters are often involved in rating the and interpreting the models. It is this qualitative element that can inform data mining techniques so that they are diagnostic in nature.

11.4 From Temporal Data Mining Techniques to Sequential Pattern Recognition

Reimann et al. demonstrated the power of top-down and bottom-up design approaches to data mining techniques to examine the temporal nature of individual and group learning, to see when and where change occurred. Jeong et al. examine the sequences in discourse patterns that lead to higher levels of argumentation. They point out that frequency data regarding specific types of discourse is important but not sufficient and propose a way to examine whether or not speech acts are predictive of more sophisticated argumentation in CSCL. In an attempt to document the evolution of an emerging argument they have developed a sequential analysis approach developed by Jeong (2005) that incorporates a coding scheme developed by Clark and Sampson (2007, 2008) that identifies, visualizes and assesses the argumentation processes in an online science environment. Sequential analysis (Bakeman and Gottman 1997) is an approach designed to analyze and model sequential links between behavioral events to determine how likely one given event is followed by another given event. Jeong et al. applied this sequential approach to argumentation and combines transitional probabilities and state diagrams to determine how students respond to specific discourse moves. In particular they identify the type of discourse moves, i.e., rebuttals, agreement, no response as well as whether the observed response patterns produces higher quality arguments, moving from claims to challenges, explanations or amendments. Proposals are made, supported or critiqued and ideas are refined. Jeong et al. propose ways to look at the discourse patterns to see which response patterns or sequences lead to changes in the argument.

Jeong et al. provide us with a concrete example of their methodology using an activity within the Web-based Inquiry Science Environment (WISE) on thermodynamics (see http://wise.berkeley.edu; Clark 2004) where individuals collect data, explore simulations and discuss observations asynchronously. The environment provides pull down menus to facilitate making principles and explaining them with others. Student discussions are facilitated by linking groups of students with other students who have made different principals then their own. This forces crosstalk and argumentation around principles that were created around a common activity where level of opposition can be coded. This structured crosstalk is also supported by a seeded discussion toolkit. Their study explores the influence of this technology in two conditions, condition one includes personally-seeded discussions based on the explanations individuals provided during their interactions, and condition two, the augmented preset condition, was augmented by researchers seeding the discussion with an optimized range of student misconceptions about principals. Students in the augmented condition demonstrated significant gains in their explanations, demonstrated more grounds for their comments, and had higher rebuttals than the other condition.

This mixed methods design is ambitious. The qualitative analysis is thorough. Discourse is coded by individual moves, examining the grounds of a comment, as well as parsing and scoring discourse episodes based on the level of opposition within episodes. Subsequent to this analysis the sequential analysis is used to identify response patterns measured in terms of the probabilities in which patterns elicit certain types of comments. The level of opposition coding is used to determine quality of argumentation and it is based on the premise that one must rebut an argument to score higher on the overall quality of argumentation. My concern with the level of opposition coding is that it does not take individual differences into account. For example, when working with high functioning groups the first argument may be quite solid and rebuttal would be unfounded. Not all groups need to rebut. In fact, it might be more intelligent to support an existing argument when it is correct than to rebut for the sake of argument. Coding schemes would need to take such individual differences into account when documenting group changes.

Jeong et al. provide an excellent mixed method approach, using a core coding scheme in addition to sequential analysis. The sequential analysis was used in a diagnostic manner demonstrating the types of patterns used in the two conditions. Their approach has great potential as a process framework for measuring discourse and argumentation in CSCL environments. It combines qualitative analysis with quantitative data mining techniques which can ultimately lead to better understanding of learning and performance. This type of research is very time consuming and there are no shortcuts. However, the careful documentation of codes of interest followed by sequential analysis can help isolate which discourse and argumentation patterns lead to higher levels of conceptual understanding. It could be used to look at the influence of teachers or tutors in promoting conceptual change as well.

11.5 Where Do We Go from Here?

The contributors to this volume have provided us with valuable insights into how we can examine and explain the role of individual learning and group processes in CSCL. It was extremely helpful to have a balanced set of articles that dealt with concrete examples using both qualitative and quantitative methods. The article by Janssen et al. was particularly beneficial in identifying some of the statistical concerns that CSCL researchers should be aware of when designing their experiments and analyses. In particular, they describe issues of hierarchical nesting, non-independence of dependent variables and differing units of analysis. They propose MLA as an answer to these three concerns and Stylianou-Georgiou et al. demonstrate the effectiveness of the approach for isolating distinct contributions of the individual and the group in their research in using hypertext to support project based learning in science.

For those researchers who do not have the benefit of large sample sizes, there were ample examples of mixed methods approaches that combine complex qualitative analyses with quantitative measures. The articles by Reimann et al. and Jeong et al. provide researchers with some alternative approaches that capitalize on mixed methods approaches. I applaud Reimann et al. for their theory-driven approach to data mining and we should recognize the difficulties they reveal in finding an appropriate theory for CSCL designs. They aptly caution against the garbage-in garbage-out phenomenon of going into an analysis blind without any preconceptions of what patterns may mean. Jeong et al. also provide a robust qualitative coding scheme for examining patterns in argumentation, which they use to guide their predictions of what discourse episodes lead to higher levels of argumentation. These combination approaches are very powerful.

In closing, these contributions have provided researchers with some gems to consider. Unfortunately, in order to get to the gem, you have to pound at the rock, requiring much effort to reveal the beauty below. CSCL researchers do not have an easy task in revealing the contributions of both the individual and the group to performance. However, mining these complex learning situations will result in better understanding of the mechanisms of learning, and will begin to operationalize the contributions of the tools, the task, the individual and the group in the learning activity in question.

References

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

Clancey, W. (1997). *Situated cognition: On human knowledge and computer representations*. Cambridge: Cambridge University Press.

Clark, D. B. (2004). Hands-on investigation in Internet environments: Teaching thermal equilibrium. In M. C. Linn, E. A. Davis, & P. Bell (Eds.), *Internet environments for science education* (pp. 175–200). Mahwah: Lawrence Erlbaum.

Clark, D. B., & Sampson, V. D. (2007). Personally-seeded discussions to scaffold online argumentation. *International Journal of Science Education, 29*(3), 253–277.

Clark, D. B., & Sampson, V. (2008). Assessing dialogic argumentation in online environments to relate structure, grounds, and conceptual quality. *Journal of Research in Science Teaching, 45*(3), 293–321.

Engeström, Y. (1987). *Learning by expanding: An activity-theoretical approach to developmental research*. Helsinki: Orienta-Konsultit Oy.

Ericsson, K. A., Krampe, R Th, & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review, 100*(3), 363–406.

Greeno, J. (1998). The situativity of knowing, learning, and research. *The American Psychologist, 53*(1), 5–26.

Jeong, A. (2005). A guide to analyzing message-response sequences and group interaction patterns in computer-mediated communication. *Distance Education, 26*(3), 367–383. Retrieved December 2008 http://garnet.fsu.edu/~ajeong/research/Jeong2005_AnalyzingInteraction.pdf.

Kay, J., Maisonneuve, N., Yacef, K., & Reimann, P. (2006). The Big Five and visualisations for team work activity. In M. Ikeda, K. D. Ashley, & T.-W. Chan (Eds.), *Proceedings of intelligent tutoring systems (ITS06)* (pp. 197–206). Heidelberg: Springer.

Kenny, D. A., Mannetti, L., Pierro, A., Livi, S., & Kashy, D. A. (2002). The statistical analysis of data from small groups. *Journal of Personality and Social Psychology, 83*, 126–137.

Mokhtari, K., & Reichard, C. A. (2002). Assessing students' metacognitive awareness of reading strategies. *Journal of Education & Psychology, 94*(2), 249–259.

Salas, E., Sims, D. E., & Burke, C. S. (2005). Is there a "Big Five" in teamwork? *Small Group Research, 36*(5), 555–599.

Strijbos, J.-W., & Fischer, F. (2007). Methodological challenges for collaborative learning research. *Learning and Instruction, 17*, 389–393.

von Glasersfeld, E. (1995). A constructivist approach to teaching. In L. Steffe & J. Gale (Eds.), *Constructivism in education* (pp. 3–16). Hillsdale: Erlbaum.

Part III
Frameworks for Analyzing Interactions in CSCL

Chapter 12
Quantifying Qualities in Collaborative Knowledge Construction: The Analysis of Online Discussions

Karsten Stegmann and Frank Fischer

Abstract Several approaches in CSCL research conceptualize and empirically explore the relations between specific qualities of text-based knowledge building processes in CSCL and successful knowledge construction. In argumentative knowledge construction, for example, a positive relation is assumed to exist between argumentation and individual knowledge construction. In order to test such assumptions and provide findings that are generalizable, the respective qualities must be appropriately quantified. This chapter describes MAQCOD (Multidimensional Approach for the Qualitative Coding of Online Discussions) which allows the theoretical assumptions of relations between qualities of text-based knowledge building processes and knowledge construction to be tested via the development of rules for segmentation and coding and the application of high-level data analysis to already coded data.

Over the last few years, a variety of new technologies for text-based communication on the internet have evolved. In addition to private communication via e-mail, public communication tools have also been developed and allow multiple possibilities for participation. Newsgroups and discussion boards offer space for the discussion of topics and for requests for support. Users write (micro-) blogs and exhibit these as a form of public diary. The blogger discusses the entries with blog visitors. Wiki users collaboratively write articles, with several authors revising the same article and discussing problematic content-related issues in corresponding discussion boards. The asynchronous collaborative writing of wiki pages may differ substantially from commenting a blog entry with respect to the knowledge building processes involved.

Process analyses are, however, not necessarily qualitative. Based on the assumption that active participation in CSCL may constitute a main factor for learning, a rather simple approach for process analysis of CSCL is to focus upon activity in general. In text-based scenarios, for example, counting the number of words or contributions made is an easy approach, which found extensive application in the pioneering phases of CSCL research (e.g., Harasim 1993). More differentiated

K. Stegmann (✉)
University of Munich, Leopoldstrasse 13, 80802 München, Germany
e-mail: karsten.stegmann@psy.lmu.de

S. Puntambekar et al. (eds.), *Analyzing Interactions in CSCL*, Computer-Supported Collaborative Learning Series, DOI 10.1007/978-1-4419-7710-6_12,

approaches usually take multiple and complex features of the discourse data into account. Approaches focusing on socio-cognitive conflicts (see Doise and Mugny 1984), for example, examine features such as the type of conflict resolution (e.g., Nastasi and Clements 1992). In these analyses, conflicts must first be identified before resolutions can be classified. In the case of cognitive elaboration – to take another example – researchers could mainly focus on the number and types of inferences that can be identified in the collaborative learning process (e.g., van Boxtel et al. 2000). This entails identifying inferences (e.g., relations created by the learner between theoretical concepts and problem information) with regard to the provided learning material and resources. These examples show that current state-of-the-art approaches in CSCL research have moved on from simple "coding and counting" methods and aim to quantify certain qualities of collaborative learning processes.

The quality of collaborative learning processes is usually defined theoretically. CSCL researchers typically assume positive relations between specific features of the collaborative process and successful collaborative knowledge construction. Criteria relating to communication norms and principles, such as the completeness of an analysis or the soundness of an argument (e.g., Spada et al. 2005; Stegmann et al. 2007b), are additionally regarded as appropriate benchmarks in defining "quality".

The present chapter presents a method for the qualitative analysis of text-based, online discussions that allows for statistical analyses and in turn a generalization of results. While the approach has been successfully applied in a number of studies (e.g. Stegmann et al. 2007a; Stegmann et al. 2007b; Weinberger 2003; Weinberger et al. 2005a) conducted at the University of Munich, it is obvious that it certainly is not the only method which can be employed for such analyses. We refer to our approach as the *M*ultidimensional *A*pproach for the *Q*ualitative *C*oding of *O*nline *D*iscussions (MAQCOD). We will suggest a somewhat idealized path from the research questions under investigation to the answers to these questions. The goal is to describe the stages involved in the development of accurate, precise, objective, reliable, reusable, and valid instruments for the analysis of processes of online discussions in CSCL. MAQCOD builds on and explicates state-of-the-art in methodology in cognitive research on learning and instruction in order to provide a kind of blueprint and point of reference for the multidisciplinary field of CSCL, where a broad variety of methodologies for dialog analysis can be found (Hmelo-Silver and Bromme 2007). Further, The chapter refers to new approaches in the field of computer linguistics which are presented in other contributions of this volume and which might help support (semi-)automated applications of MAQCOD.

12.1 Multidimensional Approach for the Qualitative Coding of Online Discussions (MAQCOD)

The respective *research question* forms the starting point of our approach. Scientific investigations begin either with a problem in educational practice or a psychological or educational theory. The process of moving from a problem or a theory to an empirical research question is beyond the scope of this chapter. Furthermore, between the research question and data analyses, data collection is usually necessary. This chapter, however, focuses on analyses and not on the gathering of data.

The *definition of data structure and the derivation of coding rules* are both strongly influenced by the research question at hand. As we will discuss later, the research question affects the unit of analysis and the categories to be differentiated. Coding rules are defined according to the unit of analysis and the required categories against the background of the respective theory.

Training of the coders then follows. We will argue that this phase may lead to changes in the data structure and the rules for coding. After finalizing the coding scheme and the corresponding coding manual, the final step of the training is to measure the degree to which different coders converge with regard to the assignment of the units of analysis (segments) to the different categories (codes).

Finally, the initial research question can be addressed using *aggregation and high-level analyses*. We will discuss the "simple" aggregation of numbers or shares of specific categories as well as examples for the computation of scores that allow for a deeper understanding of intra-group processes (such as the quantitative analysis of argument sequences or scores quantifying mutual influences within a group of learners).

12.1.1 The Initial Research Question

The first step in our heuristic is to begin with a clearly formulated *research question* and a set of (independent and dependent) variables. The main rational in developing a framework for the analysis of text-based knowledge building processes in CSCL is the research question to be addressed. CSCL researchers may, for example, formulate rather exploratory questions such as "What kind of communication will occur between students within asynchronous web-based conferencing?" (cf. Järvelä and Häkkinen 2002) where "kind of communication" specifically refers to a theory of perspective taking, or rather confirmatory questions such as "To what extent does an epistemic collaboration script affect the epistemic quality of the collaborative knowledge construction process in CSCL environments?" (e.g., Weinberger 2003). With respect to the analysis of text-based discussions, the difference between explorative and confirmative research questions mainly affects the next step in which the data structure, including the dimensions, categories, and segmentation, is defined.

12.1.1.1 Illustrative Example

Before proceeding, we will introduce an illustrative example in order to illustrate the different stages of MAQCOD. This example stems from a learning environment that was used for several experiments examining the processes during online discussions (e.g., Stegmann et al. 2007a; Stegmann et al. 2007b; Weinberger et al. 2005a). Learners had the task of analyzing three different problem cases using Weiner's attribution theory in groups of three. During an 80-min collaborative-learning phase, learners were required to discuss the problem cases in light of the theory and to find joint solutions for each of the cases. One problem case, for example, describes the situation of a school student called Michael who has bad grades in mathematics. His parents tell him that they, and as a result he too, are not particularly gifted in mathematics. His teacher tells him that he is lazy and that he could get better grades if he put in more effort.

You can find an example of one typical thread from these discussions in Figs. 12.1 and 12.2. The experimental variations employed usually involved different computer-supported collaboration scripts, that is, instructions on how to perform and sequence specific activities (cf. Weinberger et al. 2007b). The analyses in the example experiments focus, among others, epistemic activities and quality of argumentation (Weinberger and Fischer 2006).

Argumentation: First analysis *(by Maple, 6.12.2002, 10:18)*
↳***Counter-argumentation*** *(by Pine, 6.12.2002, 10:34)*
↳***Integration*** *(by Maple, 6.12.2002, 10:49)*

Argumentation: First analysis *(by Maple, 6.12.2002, 10:18)*

1. Hypotheses:
It's striking how huge the disadvantage is if someone ascribe failure to a lack of talent. Mrs. Weber goes the extra mile by her proclamation of blatant laziness. My really cool hypotheses: This guy Michael, he is and will ever be a lame duck regarding math.

Grounds:
Michael will further adhere to the belief that he is not gifted. This belief supported by his father's beliefs, Michaels failure, and his funny math teacher. Michael's lack of talent regarding math seems to be static, because "the Peters family" were a scare of all math teachers.

Qualifications:
Effort is internal variable. Michael is "curious" with respect to the next school term. Hence, he also considers feeling of success. And of course, not each failure that is ascribed to a lack of talent leads to a total stop of effort. However, there is a huge risk in the particular case of Michael. Maybe he should make a reattributional training?

Fig. 12.1 Example thread from an online discussion supported by collaboration script that provided the prompts "Hypotheses", "Grounds", and "Qualifications". Names of learners are made anonymous by using Maple and Pine instead of real names (part 1)

↳ ***Counter-argumentation*** *(by Pine, 6.12.2002, 10:34)*

1. Hypotheses:
Michael may become motivated by the desire to break through the chain of losers in his family. Maybe he's going to to get one over on his teacher. Thereby, he would ascribe his failure to an internal variable factor.

Grounds:
Students don't like teachers and aiming often to outperform their parents.

Qualifications:
Mostly for students at the same age as Michael prefer do nothing under the pretext of incomptence

↳ ***Integration*** *(by Maple, 6.12.2002, 10:49)*

1. Hypotheses:
Michael does not aim to break through a chain of losers.

Grounds:
Daddy and mommy guard his back and were not the sharpest tool in the shed either. Furthermore, his "silent purring" is an evidence for alcoholism. There it is over with breaking through chains of losers.

Qualifications:
Michael may also discover that his parents do not complying his idea of good anymore and that they are no math idols. It may work if he realized that his parent as losers AND aim not to become a loser himself. If the talent does not hinder him.

Fig. 12.2 (continued) Example thread from an online discussion supported by collaboration script that provided the prompts "Hypotheses", "Grounds", and "Qualifications". Names of learners are made anonymous by using Maple and Pine instead of real names (part 2)

According to Fischer and colleagues (Fischer et al. 2002), epistemic activities describe to what extend learners elaborate the conceptual space, the problem space, or relations between theory space and problem space in a knowledge building task. Hence, epistemic activities describe how learners work on a given knowledge building task in a systematic way within a specific domain. With regard to the scenario described above, three main types of epistemic activities are differentiated: (a) elaboration of information relating to the problem case (problem space), (b) elaboration of theoretical concepts (conceptual space), and (c) relating relevant concepts from theories and pertinent information from the problem case. With respect to the epistemic activities, an example of a research question in our experiments was "To what extent are specific epistemic activities (e.g., relating conceptual and problem space) associated with improved domain knowledge among the participating individuals?" (cf. Weinberger et al. 2005a).

The *quality of argumentation* describes the completeness of (i) single arguments and (ii) argument sequences with respect to specific models of argu-mentation (Leitão 2001; Toulmin 1958). According to Toulmin (1958), *single arguments* may consist of a claim, grounds, and qualifications. Against this background, single arguments can be differentiated into (a) bare claims, (b) grounded claims, (c) qualified claims, and (d) grounded and qualified claims. With respect to the quality of single arguments an example of a research question in our experiments is "To what extent is the completeness of single arguments related to depth of cognitive processing?" (see Stegmann et al. 2007a) According to Leitão (2001), the following main building blocks of *argumentation sequences* can be differentiated in the context of knowledge building: (a) argumentations, (b) counter-argumentations, and (c) integrations. With respect to the quality of argumentation sequences, an example of a research question in our experiments is "To what extent are longer argumentation sequences (including counter-arguments and possible integrations in contrast to just an argument) related to improved domain knowledge among the participating individuals?" (cf. Stegmann et al. 2007b).

12.1.2 Definition of Data Structure and Rules for Coding

In defining *data structure*, both top-down and bottom-up directions must be taken into account. In MAQCOD, the data structure can be described using the components *dimensions*, *categories* within each dimension, and the *size of the segments* to which the categories are applied. Typically, *dimensions* are derived from theory and constitute components of the research question (e.g., "Quality of Communication", Järvelä and Häkkinen 2002; "epistemic activity", Weinberger 2003). However, there may be cases in which dimensions can be split into several sub-dimensions; quality of argumentation might, for example, be divided into the sub-dimensions quality of single arguments and quality of argumentation sequences (cf. Stegmann et al. 2007b). This distinction may make sense if the different aspects vary rather independently. In the case of quality of argumentation, for instance, a counter-argument (sub-dimension quality of argumentation sequences) may or may not contain grounds (sub-dimension quality of single arguments).

Defining categories in a top-down manner entails a derivation of categories from theory (see running example), while bottom-up categories are developed by exploring and interpreting the data. Whether the top-down approach or the bottom-up approach is weighted more strongly or the two are balanced depends on the extent to which effects and relations between the variables under examination have already been clearly conceptualized in theory or whether they are rather to be explored in the empirical analysis. In addressing the question concerning the extent to which poor and good learners differ with respect to their contributions to an online discussion, we might prefer to emphasize the bottom-up direction. If, however, we are going to test the hypothesis that inferring relations between theoretical concepts and information from a given problem case is positively related to knowledge construction, we might instead emphasise the top-down direction. In our view, definition of the data structure is always a mixture of the two. To conduct analyses that add new findings to an existing body of research, it is necessary to build on existing concepts and methods. At the same time, valid process analyses require the integration of a rich set of features from the data source, since many relevant processes in CSCL only become visible when context information is also regarded (cf. Stahl 2006). Contextual variables are often not conceptualized by the theory of learning and knowledge building. In addition to top-down categories, some categories are therefore formulated to describe regularities in the data that cannot be accounted for using existing concepts.

12.1.2.1 Categories and Scaling

The *categories* within each dimension are either nominal or ordinal/interval-scaled. *Nominal categories* are usually applied to smaller segments such as sentences. Each nominal category represents a specific feature of the unit of analysis (e.g., grounded claim or counter-argument). The rules for coding must explicitly state when to apply which category (cf. Weinberger and Fischer 2006). After coding, the number of specific segments is aggregated for larger units such as messages or whole discussions. Due to the usually small segment size, nominal categories allow for multiple foci during analysis. Nominal coding can be aggregated across different units of analysis, for example, single messages or single learners (Weinberger and Fischer 2006). These units can be compared with respect to the number or proportion of specific categories which occur. For example, an aggregation on the level of messages (or phases; cf. Kapur et al. 2010) allows examining whether the quality of messages varies over time. With the very same data, but an aggregation on the level of single learners, the quality of discourse of different learners can be compared. Due to the multiple measurement of the feature, this approach is particularly beneficial for the generalization of results. Aggregation across several units of analysis, measurement error of aggregated measure is theoretically equal to zero (cf. classical test theory; see Novick 1966). Therefore, reliability and in turn the generalizability of findings is increased. Nominal coding also allows (due to the usually small size of segments) for the synchronisation of different data sources using a timeline (e.g., Stegmann et al. 2007b) or the computation of additional scores for the measurement of features of discussions that lack the well-elaborated theoretical basis which is necessary for an a priori definition of quality levels (e.g., knowledge convergence; Jeong and Chi 2007; Stegmann et al. 2007a).

Ordinal/interval-scaled categories usually describe the quality of a larger segment (e.g., whole discussions; Spada et al. 2005). Each category represents a specific quality of the unit of analysis with respect to a specific dimension (e.g., quality of single arguments). Whether the scale can be regarded as ordinal or interval, that is, whether the distance between different levels is equal or not, must be determined based on the theoretical assumptions and the data at hand. For both ordinal and interval scales, the coding rules must explicitly state in which cases which quality level should be coded. This requires a clear theoretical foundation as well as clearly interpretable features in the data source. For example, coding the quality level of argumentation sequences requires, at the beginning, the definition of the lowest and the highest level. Between the two extreme levels, a number of reasonable (and recognizable) levels and corresponding rules must be defined. According to Leitão (2001), a discussion at the lowest level contains no counter-argument or integration. The highest level may be coded if each argumentation is followed by a sequence of several counter-argumentations and integrations (cf. Leitão 2001). Once such rules have been defined, this procedure is faster to apply than the coding of nominal categories, since the lower number of segments requires the coder to make fewer decisions.

Defining which categories are to be coded or which quality levels are to be distinguished is not possible based solely on theoretical considerations or solely on an exploration of the data. Our suggested heuristic is to create a decision tree (cf. Fig. 12.3) that begins in a top-down manner and defines the general area of interest. The rectangles contain questions and the lines on the right hand side of each rectangle represent the complete set of (defined) possible answers. Each specific answer leads either to a subsequent question or to the final decision (i.e., the category to be coded). In the figure above, the outcomes of the decision process (i.e., the end nodes) are represented as black triangles. In our approach, we further differentiate between top-down decisions (white rectangles) and bottom-up decisions (black rectangles). In order to code a specific segment, a coder is required to make a series of decisions beginning on the left-hand side until he or she reaches the end node.

Our example for epistemic activities starts with theory-driven decisions (top-down). The coder may have to code the segment "Michael attributes his failures." Regarding theory, a problem space (e.g., the case information "Michael") and a conceptual space (e.g., the theoretical concept from the attribution theory "attribution of failure/success") can be distinguished. Attribution theory as the conceptual component of the learning environments is not to be confused with the researcher's theory that drives the top-down approach. With regard to the assumptions surrounding epistemic activities, a central task in case-based learning can be seen to constitute the construction of relations between problem space and theoretical concepts (cf. Weinberger 2003). Therefore, theory driven questions such as "does the segment contain information from the problem space?" can be applied to each scenario in which problem cases are used (see white rectangles). With regard to our example, the question "Does the segment contain information from problem space?" as well as "Is the theoretical concept explicitly related to the problem space?" can be answered with "yes". However, to enable a deeper understanding of the content of the segments, features from the data at hand must be integrated. In the case of the example provided in Fig. 12.3, decisions have to be made on the basis of the content of the problem cases (i.e., bottom-up; see grey rectangles). The cases provide information on different subjects (Michael, Michael's parents, teacher and Michael's). In combination with the concepts from attribution theory (represented in Fig. 12.3 by the letters A to F; it should be noted that attribution theory actually comprises several more concepts, while only the mentioned concepts were included in the learning material provided), a set of possible adequate relations between the problem space and the conceptual space can be defined (indicated as "R1"- "R18") and identified. In terms of the decision tree (Fig. 12.3), the sentence "Michael attributes his failures" should be coded as "R4", since the subject mentioned is "Michael" and the concept mentioned is "attributes failure".

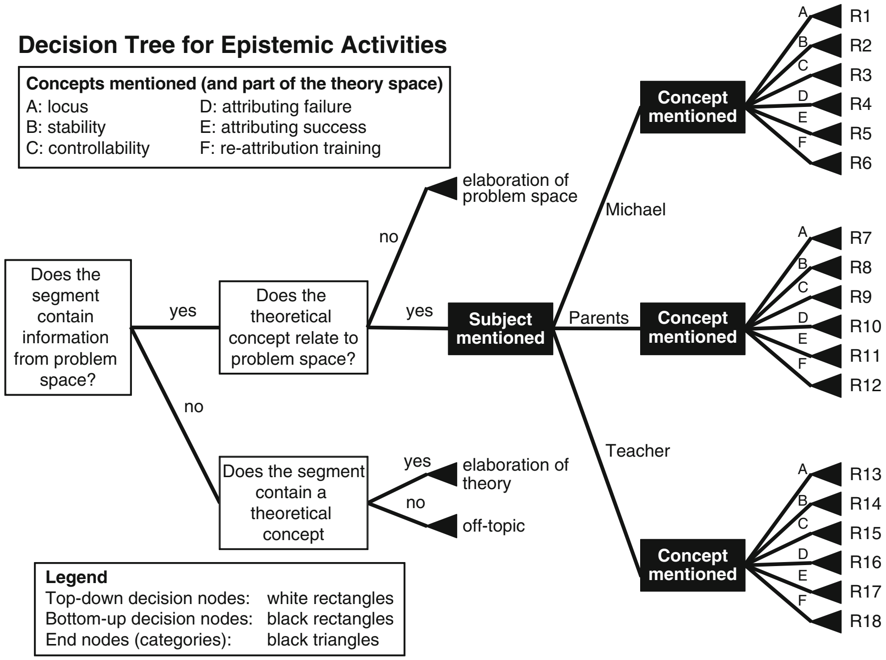

Fig. 12.3 Decision tree for epistemic activities according to Weinberger and Fischer (2006). Legend: top-down decision nodes: *white rectangles*; bottom-up decision nodes: *black rectangles*; end nodes (categories): *black triangles*

12.1.2.2 Additional Dimensions Addressing Methodological Questions

In addition to the dimensions addressed in the research question, researchers may have to include dimensions addressing methodological issues. A dimension may, for example, measure compliance to the instructions (e.g., whether learners followed the instruction to use the prompts provided to write their messages in an online discussion) in order to ensure internal validity of an experiment (i.e., conduct a manipulation check) (e.g., Weinberger 2003).

Figure 12.4 depicts an example of a previously employed collaboration script which provided the prompts “claim:”, “ground(s):”, and “qualification(s):” to support the construction of single arguments during an online discussion (cf. Stegmann et al. 2007b). All learner-generated text was coded as “no prompt”. The use of prompts was coded with respect to the learner-generated text which followed the prompt. In the example below, the prompt “claim:” is followed by a learner’s statement that actually could be classified as “claim”. This prompt would therefore be categorised as “correct use”. The prompt “ground(s):” was not used, that is, no text was typed in by the learner following the prompt. The category “ignored” would

Analyzing Compliance to the Instructions (Manipulation Check)

Message text including prompts	Use of prompt
Claim:	correct use
Michael's attribution is internally stable	no prompt
Ground(s):	ignored
Qualification(s):	non-intended use
What do you think?	no prompt

Fig. 12.4 Example of the use of categorisation in conducting a manipulation check (cf. Stegmann et al. 2007b). All learner-generated text was coded as "no prompt". The use of prompts was coded with respect to the learner-generated text which followed the prompt

therefore be applied. The category "non-intended use [of prompt]" would be applied for the prompt "qualification(s):", because instead of giving a qualification, the learner asked for feedback.

12.1.3 Segmentation

Defining the size of segments goes hand in hand with the defining of dimensions and categories. Findings at the level of messages cannot be directly transferred to the level of, for example, inferences or sentences. In the case of research on collaboration scripts (cf. Kollar et al. 2006), for instance, assumptions are made that a higher number or higher share of elaboration activities in online discussions is positively related to improved collaborative knowledge construction. If a specific feature, such as the number of grounded claims, is assumed to be positively related to knowledge construction, then determining general scores for the quality of single arguments would not be a direct way to test the assumption. Decisions regarding segmentation thus have to be made depending on the research question at hand and the respective theoretical background.

A rather extensive segmentation procedure involves the definition of segmentation rules for each dimension to be coded. Training and conducting of segmentation must be repeated for each dimension. In the case of argumentative knowledge construction, for example, the quality of single arguments is assumed to be positively related to knowledge construction. Subject to investigation is thus the number of single arguments with a high quality and segments should be single arguments (e.g., Stegmann et al. 2007b; Weinberger and Fischer 2006). In the case of cognitive elaboration, it is assumed that the application of concepts to problem information is positively related to knowledge construction and adequate segments are therefore inferences (e.g., Weinberger and Fischer 2006).

A further problem that might occur during segmentation is that the components of a single segment are not necessarily continuously written within a single sentence or section (cf. Van Dijk and Kintsch 1983). In the case of inferences, texts may even have to be re-arranged and re-written. As shown in Fig. 12.5, the information taken from the problem case ("The teacher attributes Michael's failure…") is used to create two segments on the dimension of epistemic activities, since two different theoretical concepts (internal attribution and variable attribution) are applied to the very same information. Furthermore, the sentence "She argues that Michael is plain lazy" can be regarded as a ground for both claims. In these cases, specific parts of a sentence (e.g., the subject) or whole sentences (e.g., the ground) may even be part of several segments (see Fig. 12.5, rectangles with dashed borders).

On the one hand, dimension-dependent segmentation increases the external and internal validity of the findings, since the findings are on the most adequate grain size. On the other hand, dimension-dependent segmentation requires a separate segmentation process for each dimension, including training (see section Training of coders) of the segmentation rules. This procedure clearly may lead to an enormous increase in effort. To limit the effort involved, researchers may start with the finest grained segmentation (see Fig. 12.5). All rougher grains may be an aggregation of the segments of more finely grained segments, i.e. bigger segments comprise one or more of the more finely grained segments. For example, the inferences on the dimension of epistemic activities (Weinberger et al. 2005a) are the most fine-grained (Fig. 12.5; rectangles with solid border). The next level of fine-graininess may be single arguments (e.g., bare claim, warranted claim, qualified claim; cf. Stegmann et al. 2007a) (Fig. 12.5; rectangles with dashed border) which may in turn constitute the building blocks of argumentation sequences (Stegmann et al. 2007b) (Fig. 12.5; rectangles with dotted border). This is also a reuse of the segment borders (i.e., the borders used for finer grains are reused; borders that do not already exist for all finer grains are not allowed) of the smallest units and should increase consensus among coders by reducing the number of potential places for segment borders for other dimensions.

In the case of a sentence, a segment border can be placed between each word, as a result of which, the number of the potential places for a segment border is the number of words minus one. By re-using the segment borders of a previous segmentation, the number of potential places for segment borders is reduced from the "number of words minus one" to the "number of segments minus one". The complexity of defining segments is thus reduced due to a reduction in choice and an ensuing increase in objectivity. Note, however, that objectivity of the segmentation must be independently evaluated at each level. To avoid the multiplication of errors, coders should use their own segmentation at a lower level rather than the segmentation performed by another coder. The use of segmentation conducted by another coder involves high-level errors that occur due to errors at a lower level by the other coder, but not of the current coder.

Defining dimension-dependent segments is, however, also regarded as problematic. Strijbos et al. (2006), for example, suggest that such a procedure may result in a lack of reliability with respect to segment borders. For instance, with respect to the epistemic dimension, the sentence "Michael's mother and father attribute [the failure] in an internally stable manner" leaves several degrees of freedom concerning where the border between two segments should be placed. With regard to the decision tree in Fig. 12.3, we might need to distinguish between the inferences: mother – internal attribution, mother – stable attribution, father – internal

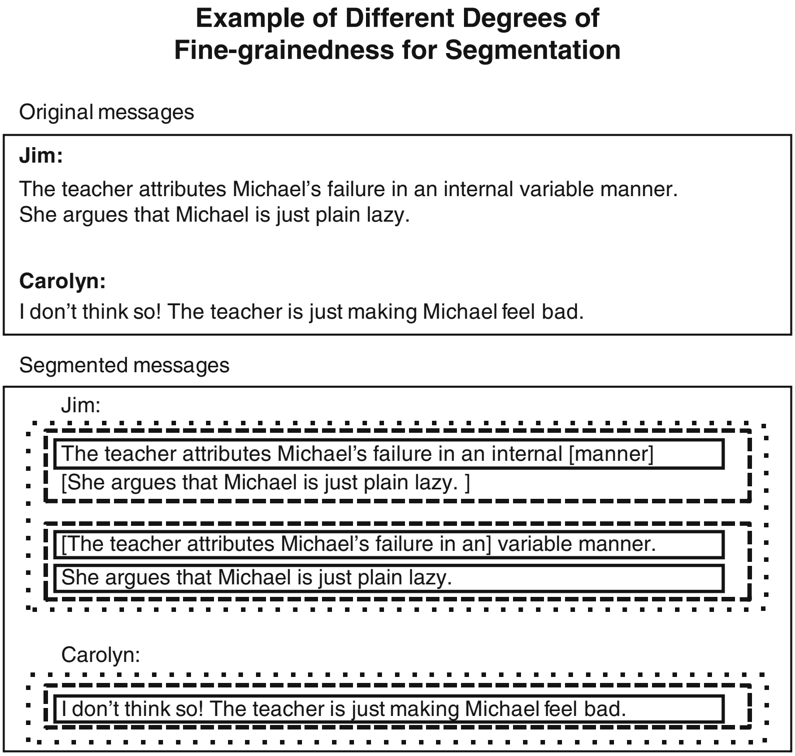

Fig. 12.5 Example of different degrees of fine-graininess for segmentation; meaning of rectangles in the segmented message: solid border=epistemic activities; dashed border=single arguments; dotted border=argumentation sequence; text in squared brackets is re-written to create meaningful statements

attribution, and father – stable attribution. But where exactly should the borders be placed? And should we really distinguish between mother and father, or should we simply regard "mother and father" as parents (i.e., two inferences instead of four)? If we do not regard them as "parents", then how should we treat the sentence "The parents attribute in an internally stable manner"? These kind of questions require defined answers in a coding manual. Ensuring agreement of coders regarding dimension-dependent segments and thus the reliability of segmentation requires rather extensive training without which it will remain suboptimal. Strijbos et al. (2006) therefore suggest segments that are rather independent of categories but that can be identified with a high degree of accuracy by different coders. They exemplify their approach by presenting clear segmentation rules for email and propose a segmentation of compound sentences for this type of text-based knowledge building processes in CSCL. However, this procedure is only appropriate if the unit from which the researcher aims to draw inferences is not considerably smaller (e.g., propositions) or considerably larger (e.g., messages) than compound sentences. In all other cases, the validity of the analysis is endangered.

12.1.3.1 Skipping Segmentation

In some cases, the unit from which the researcher aims to draw inferences from has natural borders. For example, when examining the argumentative quality of messages, whole messages might be regarded as a single segment (e.g., Marttunen 1994). In such cases, the manual segmentation process can be skipped. However, these natural borders may cause other problems. For example, in chat communication, turn taking can be regarded as a natural border of a segment although one activity may spans across several turns. A further example is that long messages in an email communication may contain indicators from more than one category although only one category can be assigned. This problem might be addressed by employing ordinal/interval-scaled categories, which enables the definition of rules for the evaluation of segments with multiple features (e.g., one counterargument and two arguments on the dimension of quality of argumentation sequences).

However, if the number of segments in which a specific category is coded is irrelevant and convergence with, for example, an expert analysis is to be examined, then segmentation of the text-based knowledge building processes can be completely skipped. Instead, the expert analysis can be divided into single items so that the coder can decide whether each item is present within the knowledge building processes. Jeong and Chi (2007), for example, created a template of 178 stated or inferred “knowledge pieces”. The two coders were to decide which of these knowledge pieces were to be found in the protocols of the students’ collaboration. With regard to our running example, we would search for the 18 relations between the problem space and the conceptual space identified as appropriate (see Fig. 12.3). Based on the decision tree, however, we would not be able to examine the role of repeating concepts from the theory or repeating information from the problem space. Therefore, a complete “search list” of theory concepts or facts from the problem case must be created. Chi (1997, p. 289) refers to this procedure as “searching rather than segmenting”. However, the limitations of this procedure are obvious: a search list must first be created and multiple occurrences of single list items are not considered (with respect to the theoretical background). For example, given the research question “to what extent is the completeness of single arguments related to the depth of cognitive processing”, the quality of single arguments can hardly be measured using a search list. Identifying one complete single argument does not allow for an assessment of the relationship between completeness of single arguments and depth of cognitive processing. Instead, the number or share of complete arguments must be measured.

12.1.4 Training of Coders

Commencing with the *training of coders* often does not mark the end of the coding-scheme development process but rather constitutes the beginning of the refinement process. The early phase of the training usually comprises several discussions that (should) lead to a similar interpretation of the rules and the data source. It is not unusual that considerable changes are made to the coding scheme with respect to both segmentation and categories during this phase. The systematic analysis of several texts allows exceptions to become evident. The handling of these exceptions must be defined. This may include extending the dimension by adding a further category or a rule stating that specific exceptions extend an already existing category. In addition to the treatment of exceptions, the coding manual must be revised to include examples and a more precise formulation of rules where necessary.

12.1.4.1 Training for High Objectivity (Without a Loss of Validity)

The central goal of the training process is to achieve high objectivity and high validity in the coding that follows the training. Training thus does not primarily aim to teach coders specific skills. Specific skills are thus only necessary as far as reaching the goal of highly objective (or inter-subjective) and valid measures are concerned. Selection of the training material (i.e., parts of the data source) thus has to follow specific rules.

The training material must be representative with respect to the whole data source. In the case of experiments, coders must be trained with data from all experimental conditions in order to ensure that they recognise specific categories with a similar degree of reliability across experimental conditions. Consequently, training materials should provide the opportunity to apply all codes to data from all experimental conditions. In research on CSCL, *a low prevalence of certain highly valued processes* (e.g., socio-cognitive conflicts and a beneficial resolution of conflicts) is often the starting point for the development of an instructional intervention, so that data from a control condition may not contain the processes that coders are looking to identify. More generally speaking, the features to be identified are often very rare (e.g., Clark and Sampson 2007; Kollar et al. 2007).

Depending on the structure of the data under examination, specific criteria have to be met by the coders before their coding can be regarded as sufficiently objective. If the data needs to be segmented, the ratio of agreement regarding segment borders must first be calculated. This can be done by identifying the number of discrepancies between two coders, that is, the number of segments that are only identified by Coder A and the number of segments that are only identified by Coder B. A more detailed discussion concerning the issue of when segments can be regarded to have been identically identified by two coders is provided by Strijbos and colleagues (2006). Furthermore, the maximum number of different segments must also be calculated, that is, the total number of segments identically identified by both coders and the number of discrepancies between the two. By dividing the number of segments identically identified by both coders by the maximum number of different segments, the ratio of inter-rater agreement for segmentation can be calculated. This ratio should generally be higher than 80% (cf. Riffe et al. 1998).

Due to the fact that segmentation strongly affects the coding process, the greater the objectivity of the segmentation, the better the subsequent training steps can be performed. If segments contain ambiguous features, which lead to lower inter-rater agreement (e.g., the segment “Michael attributes internally stable” contains features that allow two different inferences to be coded on the epistemic dimension; see Fig. 12.3), this may make training more difficult. A similar problem can occur if high agreement is due to rules that disregard the validity of the segmentation. If the segmentation rules, for example, are based on sentences, while the relevant units are arguments, then the validity of the coding process is endangered. The contribution “The teacher attributes in an internally stable manner. She argues that Michael is plain lazy” is a grounded claim if analysed as a single segment, but can be regarded as two bare claims if separately analysed as two segments. In conclusion, the task during training is to refine the rules for segmentation and to simultaneously optimise objectivity and validity.

The next step in the training process is to apply the rules from the coding manual to the data segments. Two or more coders have the task of assigning the very same categories (the very same quality levels in the case of ordinal/interval-scaled categories) to the very same segments. Since agreement will occur due to chance as well as to training, Cohen’s Kappa is often computed (other options include Fleiss’ Kappa or Krippendorff’s Alpha). Cohen’s Kappa values should not fall below .4 in order to be regarded as sufficient. Values of .75 or higher indicate high chance-controlled agreement between coders and are thus preferable (cf. Krippendorff 1980).

Initial attempts to attain sufficient inter-rater agreement often fail. The best way to improve inter-rater agreement is to analyse the contingency tables that form the basis of the Kappa value. These tables help identify categories with lower consensus following which the rules for these categories must be refined. The refined rules are subsequently applied to new raw data and a new Kappa is computed. This cycle should be repeated until a sufficient Kappa value is reached. However, some categories remain hard to code even after several training cycles. A pragmatic solution for this problem is to attempt to find a broader category that can be identified with a greater degree of accuracy. The limit for this “collapsing” of categories is exceeded when the research question can no longer be answered. Collapsing the categories “bare claim” and “grounded claim” to form one single category might, for example, mean that a research question concerning argumentation can no longer be answered.

12.1.4.2 Handling Biased Data

A specific bias with respect to the training material occurs if the experimental condition can be easily recognised by the coders during the analyses. This may lead to a systematic error, with coders being more likely to apply specific categories in specific conditions. It is therefore necessary to test the objectivity of the coding whenever data provide clear indicators of the corresponding experimental condition (e.g., specific prompts that aimed to trigger specific activities). Stegmann and colleagues (2007) have presented an approach which can be used to conduct such a test. While coders in their study were unaware of participants’ characteristics, this was not the case with respect to experimental condition, which was easily recognisable. Experimental variation was realised by prompts that were added to the beginning of messages created by a participant. The prompts were thus part of the discussions to be analysed (an example can be seen in Fig. 12.4). In testing for a potential bias in favour of the experimental condition, the impact of condition awareness on coding results was determined by comparing the coding of originally scripted discourse to that of discourse without the script prompts. 25% of all scripted discussions were manipulated in two different ways: (1) all pre-specified prompts were deleted from the scripts (condition “without prompts”) or (2) pre-specified script labels were replaced (if necessary) by argumentative conjunctions such as “because”, “therefore”, or “hence” (condition “prompts replaced”). Two coders who were not informed about the experimental design or the intervention were trained to apply the coding scheme. Only material from the control condition was used for their training. Both coders coded material without prompts and different material with replaced prompts. Stegmann and colleagues (2007) found a high degree of coder agreement for discussions with original prompts and those without and with replaced prompts. In this case, awareness did not have a strong effect on the coding procedure. In the case of low coder agreement, discussions with prompts must be manipulated in such a way as to reduce the effect of condition awareness.

12.1.5 Aggregation and High-Level Analyses

The next question is how to proceed with the coded data generated by an (multi-dimensional) analysis of text-based knowledge building processes in CSCL. To a certain extent, many approaches in CSCL assume that the number or the proportion of specific processes during collaboration is related to knowledge construction. Categories are therefore usually aggregated in order to determine the number or share of occurrences of a specific category (e.g., counter-arguments) in the online discussion per learner (or discussion). After coding, the dataset often contains multiple rows per learner (i.e., one row per segment), whereas only one row per learner remains after aggregation. Instead of information regarding which category was applied to a specific segment, the table contains information on the number and/or share of occurrences of a specific category per learner. This is, however, only the first step. Once detailed data from several online discussions have been represented within the columns and rows of a statistical software package, several parameters beyond simple means, frequencies and sums can be computed. According to the research questions of our illustrative example, the amount of relations between conceptual and problem space, the amount of grounded and qualified claims, or the probability that an argument was answered with a counterargument can be computed.

The following section will shed light on recent approaches and their findings regarding how to address central questions in CSCL research. These approaches make use of quantified qualitative methods for high-level analyses. High-level analyses are regarded as comprising the analysis of coded raw data which are aggregated to create new features. The frequency of a category thus cannot be regarded as a high-level analysis, since no new feature is formed. In contrast, using the standard deviation of category frequency within a group of three to operationalize *divergence within the group* is regarded as high-level analysis.

In the following, we will illustrate such approaches based on three exemplary high-level analyses: (1) tracing knowledge, (2) quantifying mutual influence, and (3) visualising mutual influence.

12.1.5.1 Tracing Knowledge

The question concerning the extent to which learners mutually influence one another has attracted considerable attention in CSCL research. Recently, a particular focus has been placed on the degree to which groups of learners share mutual understanding via social interaction. Accordingly, attempts have been made to quantify this process referred to as knowledge convergence, based on analyses of text-based knowledge building processes. Jeong and Chi (2007) analysed the extent to which collaborative learners shared so-called knowledge pieces either through collaboration or, in contrast, through learning in the same learning environment. As described above, knowledge pieces from an expert solution were used as a search list. The transfer of knowledge pieces from one learner to another was quantified by comparing the knowledge pieces mentioned by single learners before, during, and after collaboration. Quantitative analyses revealed that shared understanding due to collaboration was rather rare and was directly related to the amount of collaborative interaction. Furthermore, the authors examined the extent to which learners developed more similar mental models. To this end, individual knowledge tests were coded with respect to the most likely mental model underlying the content under consideration. Overall, the analyses allowed conclusions to be drawn regarding the degree of knowledge convergence that can be traced back to collaboration as well as regarding the relationship between collaborative interaction and knowledge convergence.

12.1.5.2 Quantifying Mutual Influence

Weinberger et al. (2007a) have presented an approach similar to that of Jeong and Chi (2007) but with additional quantitative measures of convergence of prior knowledge, collaborative process, and acquired knowledge based on fine-grained (i.e., segmentation at the level of inferences) analyses of text-based data sources (pre-test, text-based online discussion, post-test). The authors suggest using the standard deviation of the number of different inferences within a group of learners (analogous to knowledge pieces) as an indicator of knowledge divergence. Applying these measures provided insight into the relationship between the processes and outcomes of collaborative knowledge construction (c.f. Weinberger et al. 2005b). For example, learners with high divergence during online discussion (i.e., contributing different rather than identical inferences) are more likely to share knowledge after collaboration than learners with high convergence during online discussion.

12.1.5.3 Visualising Mutual Influence

An approach to collaborative knowledge construction that focuses on a specific aspect of the mutual influence of learners beyond content-related influences is that of argumentative knowledge construction (Jeong et al. this volume; Kollar et al. 2007; Stegmann et al. 2007b). Here, one particular question refers to the role of sequences of arguments, that is, the extent to which sequences like arguments, counter-arguments, and integration are related to knowledge construction. Stegmann and colleagues (2007) coded each message of an online discussion with respect to its role in a sequence of arguments as argument, counter-argument, or integration. Subsequently, the software tool MEPA (Erkens 1998) was used to compute the probability of specific sequences (e.g., the probability that an argument was responded to with a counter-argument, etc.). Such analyses are based on the assumption that not only the frequency of an activity but also its position within a sequence of activities is important for knowledge construction. The importance of the analysis of sequences is underlined by the findings by Jeong and colleagues (this volume).

This is, of course, not an exhaustive list of the current approaches to high-level analysis of text-based knowledge building processes in CSCL. Nonetheless, the presented examples show that research which aims to extend beyond the investigation of direct effects of instruction on individual knowledge construction in groups must include such high-level process analyses.

12.2 Automated Analyses of Online Discussions Using MAQCOD with Natural Language Processing

Versions of MAQCOD have been used for exploring the potentials of current language technologies for (semi-) automated analyses of online discussions. Developments in the field of natural language processing enable technologies to emerge which support and streamline multi-dimensional analyses of collaborative learning processes. Such developments will have a tremendous impact on this increasingly important aspect of CSCL research. A study conducted together in collaboration with computer linguists (Rosé et al. 2008; see also Gweon et al. this volume, for a similar study) we found evidence that even sophisticated coding schemes (e.g., epistemic activities or argumentation sequences) which require intensive training for human coders can be largely automated. MAQCOD still needs to be applied in a manual coding process to generate data that can be used to train algorithms capable of classifying automatically. Hence, these algorithms usually do not incorporate the coding rules as defined by the researcher. Instead, specific features of the text (like length, specific words) increase the probability of specific codes. Researchers have to be careful, because the automatic discourse analysis is vulnerable to assign codes due to superficial features (e.g., very long statements are assigned as critical or each use of "because" lead to code "grounded claim").

Beyond increased efficiency of data analysis that might speed up the research process, automatic coding technology also holds the promise of more flexible kinds of instructional interventions. For example, automatic on-line analysis in real-time chat interactions may enable instructors to monitor the progress of multiple interactions occurring in parallel, indicating where the instructor's intervention is most needed and even what specific needs the instructor should address. Looking further ahead in to the future, a fully automatic system which might be still restricted to a specific dimension or a group of categories (e.g., identifying counter-arguments) might also enable automatic adaptive interventions for collaborative learning which are more flexible than current static interventions. Such interventions might, for example, include a collaboration script for argument construction which is temporarily applied when learners do not ground and warrant their claims, or do not formulate counter-arguments at all, and which is gradually faded out when learners develop internal cognitive scripts that guide their argumentative knowledge construction. Such systems could help to prevent effects such as over-scripting (Dillenbourg 2002) or negative interaction effects between internal scripts and external collaboration scripts (Kollar et al. 2007).

12.3 Summary and Conclusion

This chapter presented MAQCOD, detailing one tested way in which qualities can be quantified in online discussion and in which qualitative findings can, to a certain extent, be generalised. We set the *research question* as a starting point for our heuristic. Both the *definition of data structure and the derived rules for coding* are strongly influenced by the research question. We discussed how the research question affects the unit of analysis. In our approach, the *training of coders* follows. Finally, *aggregation and high-level analyses* are required to answer the initial research question. We further discussed the opportunities offered by an automation of the analyses including an application of current natural language processing technologies in combination with MAQCOD.

Experimental studies that examine the effects of instructional means of support for processes of collaborative learning are the main field of application for MAQCOD. While the approach can of course also be applied to case studies, the strength of the quantification of qualities and the use of high-level analyses are ideally employed in combination with statistical tests. Nonetheless, the effort which is necessary to apply the approach in a valid and reliable way should not be underestimated. Multi-dimensional, multi-level coding schemes such as the framework presented by Weinberger and Fischer (2006) require several weeks of training before coders are able to apply the framework to raw data from online discussions. Use of the coding scheme itself is also a time-consuming process and must be carefully monitored to ensure a consistent quality of coding. However, this procedure allows to test comparatively simple hypotheses (e.g., an argument script will facilitate the number of grounded arguments and the number of counter-arguments in online discussions) as well as much more complex hypotheses that can be derived from sophisticated theoretical models of collaborative knowledge building (e.g., learners of a group converge more strongly with respect to their knowledge if they diverge in the collaboration process). To enable the statistical testing of hypotheses on complex knowledge building processes in online discussions might be seen as the major advantage of this approach: It thus can support the empirical validation of mechanisms of collaborative learning assumed to mediate between design aspects and conditions of collaborative learning environments and individual learning and knowledge advancement.

However, an automation of the coding process may potentially lead to a reduction in the effort required and increase the speed with which quantified qualitative analyses of collaborative processes are performed.

Research of this kind tends to change in quality over time. As long as humans are responsible for application of the coding schemes, there is no guarantee that the quality of coding remains constant over time. Therefore, MAQCOD must integrate procedures to control for continued coding quality after training completion.

In addition, this kind of research is time consuming and requires enormous efforts of a research team. One current hope is directed towards the possibilities of automated coding. Automation would, for example, allow for the autonomous development and training of coding schemes by a single researcher. As described in the section "Training of coders", discussing the categories with other human coders has so far led to an ongoing refinement and control of validity. With tools such as the TagHelper (Rosé et al. 2008), time consuming consensualization processes can be facilitated. Training may subsequently lead to a further refinement of rules for higher agreement with the tool and in turn increase objectivity (which, in turn, requires steps to ensure the validity of coding). MAQCOD of the near future will thus need to include proposals on how to integrate natural language processing technologies into the analysis process at the same time as ensuring the validity of coding.

References

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

Clark, D. B., & Sampson, V. D. (2007). Personally-seeded discussions to scaffold online argumentation. *International Journal of Science Education, 29*(3), 253–277.

Dillenbourg, P. (2002). Over-scripting CSCL. In P. A. Kirschner (Ed.), *Three worlds of CSCL: can we support CSCL* (pp. 61–91). Heerlen: Open University of the Netherlands.

Doise, W., & Mugny, G. (1984). *The social development of the intellect*. Oxford: Pergamon.

Erkens, G. (1998). *Multiple episode protocol analysis (MEPA 3.0)*. Utrecht: Department of Educational Sciences.

Fischer, F., Bruhn, J., Gräsel, C., & Mandl, H. (2002). Fostering collaborative knowledge construction with visualization tools. *Learning and Instruction, 12*, 213–232.

Gweon, G., Jeon, S., Lee, J., Finger, S., & Rosé, C. P. this volume. A framework for assessment of student project groups on-line and off-line. In S. Puntambekar, G. Erkens, & C. Hmelo-Silver (Eds.), *Analyzing interactions in CSCL: Methods, approaches and issues* (pp. 293–317). Berlin: Springer.

Harasim, L. (1993). Collaborating in cyberspace: using computer conferences as a group learning environment. *Interactive Learning Environments, 3*(2), 119–130.

Hmelo-Silver, C. E., & Bromme, R. (2007). Coding discussions and discussing coding: research on collaborative learning in computer-supported environments. *Learning and Instruction, 17*(4), 460–464.

Järvelä, S., & Häkkinen, P. (2002). Web-based cases in teaching and learning: the quality of discussions and a stage of perspective taking in asynchronous communication. *Interactive Learning Environments, 10*, 1–22.

Jeong, H., & Chi, M. T. H. (2007). Knowledge convergence and collaborative learning. *Instructional Science, 35*, 287–315.

Jeong, A., Clark, D. B., Sampson, V. D., & Menekse, M. this volume. Sequential analysis of scientific argumentation in asynchronous online discussion environments. In S. Puntambekar, G. Erkens, & C. Hmelo-Silver (Eds.), *Analyzing interactions in CSCL: Methods, approaches and issues* (pp. 207–233). Berlin: Springer.

Kapur, M., Voiklis, J., & Kinzer, C. K. (2010). A complexity-grounded model for the emergence of convergence in CSCL groups. In S. Puntambekar, G. Erkens, & C. Hmelo-Silver (Eds.), *Analyzing interactions in CSCL: methodologies, approaches and issues*. Berlin: Springer.

Kollar, I., Fischer, F., & Hesse, F. W. (2006). Collaboration scripts - a conceptual analysis. *Educational Psychology Review, 8*(2), 159–185.

Kollar, I., Fischer, F., & Slotta, J. D. (2007). Internal and external scripts in computer-supported collaborative inquiry learning. *Learning and Instruction, 17*(6), 708–721.

Krippendorff, K. (1980). *Content analysis: an introduction to its methodology*. Beverly Hills: Sage.

Leitão, S. (2001). *Analyzing changes in view during argumentation*: A quest for *method*. *Forum qualitative sozialforschung//forum: qualitative social research* retrieved December 15, 2003, from http://www.qualitative-research.net/fqs/fqs-eng.htm

Marttunen, M. (1994). Assessing argumentation skills among Finnish university students. *Learning and Instruction, 4*(2), 175–191.

Nastasi, B. K., & Clements, D. H. (1992). Social-cognitive behaviors and higher-order thinking in educational computer environments. *Learning and Instruction, 2*, 215–238.

Novick, M. R. (1966). The axioms and principal results of classical test theory. *Journal of Mathematical Psychology, 3*(1), 1–18.

Riffe, D., Lacy, S., & Fico, F. G. (1998). *Analyzing media messages: using quantitative content analysis in research*. Mahwah: Lawrence Erlbaum.

Rosé, C. P., Wang, Y. C., Cui, Y., Arguello, J., Stegmann, K., Weinberger, A., et al. (2008). Analyzing collaborative learning processes automatically: exploiting the advances of computational linguistics in computer-supported collaborative learning. *International Journal of Computer-Supported Collaborative Learning, 3*, 237–271.

Spada, H., Meier, A., Rummel, N., & Hauser, S. (2005). A new method to assess the quality of collaborative process in CSCL. In T. Koschmann, D. Suthers, & C. Tak-Wai (Eds.), *Proceedings of th 2005 conference on computer support for collaborative learning: learning 2005: the next 10 years!* (pp. 622–631). Taipei: International Society of the Learning Sciences.

Stahl, G. (2006). *Group cognition*. Cambridge: MIT Press.

Stegmann, K., Wecker, C., Weinberger, A., & Fischer, F. (2007). Collaborative argumentation and cognitive processing - an empirical study in a computer-supported collaborative learning environment. In C. Chinn, G. Erkens, & S. Puntambekar (Eds.), *Mice, minds, and society. CSCL 2007. The computer supported collaborative learning (CSCL) conference 2007* (Vol. 8, pp. 661–670). New Brunswick: ISLS.

Stegmann, K., Weinberger, A., & Fischer, F. (2007). Facilitating argumentative knowledge construction with computer-supported collaboration scripts. *International Journal of Computer-Supported Collaborative Learning, 2*(4), 421–447.

Strijbos, J.-W., & Fischer, F. (2007). Methodological challenges for collaborative learning research. *Learning and Instruction, 17*, 389–393.

Strijbos, J.-W., Martens, R. L., Prins, F. J., & Jochems, W. M. G. (2006). Content analysis: what are they talking about? *Computers & Education, 46*(1), 29–48.

Toulmin, S. (1958). *The uses of argument*. Cambridge: Cambridge University Press.

van Boxtel, C., van der Linden, J., & Kanselaar, G. (2000). Collaborative learning tasks and the elaboration of conceptual knowledge. *Learning and Instruction, 10*(4), 311–330.

Van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York: Academic.

Weinberger, A. (2003). *Scripts for computer supported collaborative learning effects of social and epistemic cooperation scripts on collaborative knowledge construction*. Unpublished doctoral dissertation, Ludwig-Maximilians-Universität, München.

Weinberger, A., Ertl, B., Fischer, F., & Mandl, H. (2005a). Epistemic and social scripts in computer-supported collaborative learning. *Instructional Science, 33*(1), 1–30.

Weinberger, A., & Fischer, F. (2006). A framework to analyze argumentative knowledge construction in computer-supported collaborative learning. *Computers & Education, 46*(1), 71–95.

Weinberger, A., Stegmann, K., & Fischer, F. (2005b). *Measuring knowledge convergence: Achievement similarity and shared knowledge in computer-supported collaborative learning*. Paper presented at the 11th biennial conference of the European Association for Research on Learning and Instruction (EARLI).

Weinberger, A., Stegmann, K., & Fischer, F. (2007a). Knowledge convergence in collaborative learning: concepts and assessment. *Learning and Instruction, 17*(4), 416–426.

Weinberger, A., Stegmann, K., Fischer, F., & Mandl, H. (2007b). Scripting argumentative knowledge construction in computer-supported learning environments. In F. Fischer, H. Mandl, J. Haake, & I. Kollar (Eds.), *Scripting computer-supported communication of knowledge – cognitive, computational, and educational perspectives* (pp. 191–211). New York: Springer.

Chapter 13
An Interaction-Aware Design Process for the Integration of Interaction Analysis into Mainstream CSCL Practices

Alejandra Martínez-Monés, Andreas Harrer, and Yannis Dimitriadis

Abstract The potential capabilities of computers to support analysis of interaction data have attracted the attention of the CSCL research community. This has led to the proposal of a number of *interaction analysis tools*, which process interaction data to meet different purposes. These may range from supporting researchers in ethnographic studies to providing advice to the students. However, after several years working with classroom-based CSCL experiences, we have found that both researchers and practitioners meet many difficulties to apply these potential benefits to their CSCL settings. Thus, the first goal of this chapter is to provide a systematic analysis of the problems that can be found when trying to apply interaction analysis tools to CSCL settings, which are then classified at into three levels, namely: application, architecture and design levels. Then, we outline the path for possible solutions to face these problems. According to this, the issues identified at the design level call for an IA-aware design process where we distinguish between *co-design* approaches that directly integrates the diverse needs of learning and analysis, and *multi-perspective* approaches that treat them independently at an initial stage. On the other hand, the problems at the application and architecture levels must be faced by technology-driven solutions, such as the use of decoupled architectures, either based on inter-process communication or on interchange of log file information. Several open issues have also been detected that need adequate solutions, as e.g., the semantic integration of log-files when multiple self-contained learning tools are used for an integrated analysis.

A. Martínez-Monés (✉)
GSIC-EMIC Group, University of Valladolid, Valladolid, Spain
e-mail: amartine@infor.uva.es

S. Puntambekar et al. (eds.), *Analyzing Interactions in CSCL*, Computer-Supported Collaborative Learning Series, DOI 10.1007/978-1-4419-7710-6_13,

13.1 Introduction

The massive use of computers to support learning has brought the possibility to apply e-research capabilities to the learning sciences (Markauskaite and Reimann 2008). Computers can store large amounts of interaction data that can be then analyzed by automatic or semi-automatic means to serve different purposes, from pure research to formative evaluation or monitoring approaches. This challenge is even stronger in CSCL scenarios, where the complex and multimodal interactions among participants are totally or partially mediated by computers, and thus, not directly observable by traditional means in the remote scenarios. This has raised the interest of at least two different trends: those coming from ethnographic or ethnomethodology traditions, that see the computer as a tool to help researchers store and analyze detailed accounts of the interactions (Guribye and Wasson 2002), and those coming from engineering fields aiming to produce automatic or semi-automatic results that can directly help researchers or practitioners in their work (Moreno and Ventura 2007). Tools coming from these two trends have been conceptualized in Soller et al. (2005) as a continuum from mirroring tools, that store and reproduce interaction data to facilitate their analysis, to guiding tools that perform themselves the analysis and give direct advice to their users. Due to their emphasis in the analysis of interactions among participants, we will refer to all these tools as computer-based Interaction-Analysis (IA) tools, or IA tools for short. Therefore, an IA tool in this work must be understood in a broad sense as any (software) system able to take interaction data as an input, process it, and show the results of the analysis to its users. These users may be researchers, teachers or students, depending on the specific case. The format used to display the results may be very different, depending on the target user and the purpose of the analysis.

The potential benefits of computer-based analysis of interactions among participants has led to a significant growth of interest in terms of research papers (Harrer et al. 2009), meetings (Dwyer et al. 2008; Law et al. 2009), and projects (see for example, those funded by the European Commission as *Argunaut* (De Groot et al. 2007) or *Kaleidoscope* (Kaleidoscope 2007), or other international initiatives such as several international research collaborations located in the Pittsburgh Science of Learning Center (PSLC). However, in spite of the major impact in the research community, IA tools have not found yet their place in the classroom, and they have not been incorporated in real-life CSCL scenarios beyond pilot studies directly guided by researchers (Martínez-Monés et al. 2008a).

Several reasons hinder IA tools to move into mainstream educational practices (Dimitracopoulou 2005; Soller et al. 2005). Some of them deal with the difficulty of having educators designing and using IA tools in CSCL. First, there is a general resistance to adoption of technological innovations in classroom, especially when these require a major shift in pedagogy, as it happens in collaborative learning. On the other hand, there is an intrinsic complexity of the concepts and indicators involved in interaction analysis of collaborative learning. Thus, it is really difficult to advance in the definition and selection of appropriate indicators, as well as their visualization to the different actors involved in the teaching/learning process (teachers, students, evaluators). Finally, several technological problems have been reported with their origin in a significant mismatch between the learning management systems, be them generic or specific for CSCL, and the IA tools or services. In that latter case, it is not straightforward to integrate all these tools and services in platforms and put them in practice (Markauskaite and Reimann 2008). These difficulties have been experienced by the authors in several international and local research projects, and constitute the main motivation of the reflections shared in this chapter.

We present a systematic analysis of the problems and the eventual paths to solutions related to a wider adoption of IA tools and practices in CSCL environments. The approach followed here is mainly data-driven or bottom-up, illustrating both problems and solutions with examples drawn from the mentioned research projects and classroom-based case studies in which the authors have been directly involved. Although such an approach does not guarantee any generalization, it is expected that the theoretical analysis and discussion presented in this document may foster a major reflection and eventually an approach adopted by the global community of researchers, educational practitioners and technology providers.

Section 13.2 presents a general IA process model that allows a common understanding of the field, as it emerged through various projects and the joint effort of several research teams within the European Kaleidoscope Network of Excellence. Then, Sect. 13.3 proposes a classification of problems for the integration of interaction analysis into mainstream CSCL practices, illustrated through several examples drawn from practice. The following section provides a set of solutions that can be considered as design or technology-oriented. The last section includes an overall discussion based on the main conclusions and suggests a series of orientations and future steps that may be taken into account by the global community.

13.2 Overview and Model of the Interaction Analysis Process

As mentioned beforehand, computer supported interaction analysis has raised a growing attention in the learning sciences. A good example of this interest is the fact that it was a prominent theme in the mentioned Kaleidoscope Network, where several projects and initiatives took place in order to integrate and leverage current research on computer-supported interaction analysis (Kaleidoscope 2007). However, from the beginning, it became clear that the different research perspectives that converged in these projects did not share their understanding of the involved processes, tools, and methods. Thus, a major goal of the projects was to define shared conceptualizations of the interaction analysis process. The first model was produced in the ICALTS project, and consisted on a framework to describe the main concepts underlying computer-supported interaction-analysis. This model illustrates the close interplay between context, the CSCL environment and the IA tools, and helps to conceptualize single interaction analysis methods. Later, in the

CAViCoLA project, the scope of this model was extended, by defining a general framework for the whole process of computer-supported collaboration analysis, considering aspects such as data and method triangulation, appropriate for the analysis of complex learning scenarios (Harrer et al. 2007). Their main purpose was to enable researchers from different traditions share their concepts and eventually their methods and tools to support these processes. With this objective, but on a more technical level, in the IA project we proposed a *common format* to enable interoperability among CSCL and IA tools (Harrer et al. 2009). This format will be discussed later, as it is an example of the possible solutions we propose in this chapter for a major adoption of IA tools in CSCL.

The *ICALTS* interaction analysis model is especially useful to understand the interplay between learning environments and IA tools, and will be shortly described here, as it is a useful model to frame the discussions included in this chapter.

As shown in Fig. 13.1, the analysis of participants' interactions is usually driven by some sort of hypothesis which shall be proven or rejected by certain observation. So the first question is "What are the important questions to ask?", "What do I want to analyze?". The answer to this question influences (indicated by the large arrow in Fig. 13.1) the choice of *indicators* used to conduct further analysis. The questioner will choose an indicator able to express the concept to be analyzed, i.e., the choice of an indicator is influenced by the target group and will vary with the interest of the questioner. For example, researchers might want to obtain a detailed view of the process, while students might benefit from visualizations of their participation in a forum as compared to other students in the same classroom.

The choice of an indicator determines certain constraints a *learning environment* has to fulfill (indicated by the dashed arrow in Fig. 13.1). Thus, each indicator determines "what should be available to compute the indicator's values?", e.g., to measure social structures or patterns of interactions it is necessary to capture the information

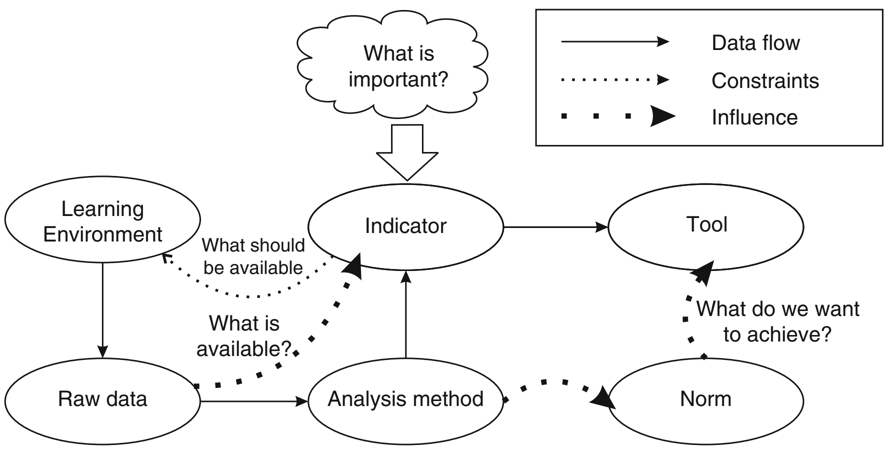

Fig. 13.1 Schema of the interaction analysis process

related to “who is sending messages to whom”, etc. The learning environment is responsible for generating the *raw data* (e.g. log files) used in further analysis steps to compute the chosen indicators. Sometimes the data required by an indicator cannot be supplied by the used learning system causing the set of indicators to be narrowed or re-assembled. Thus, it can be seen that the availability of appropriate raw data influences the choice of indicators as well (shown by the curved arrow). These mutual constraints between the learning environment and the IA tool will be illustrated in the next section with examples taken from the authors’ experience.

In the end the *analysis method* relays the indicators to a certain *tool* that uses these indicators and eventually presents them to the intended target users (researchers, teachers, students, etc.). In some circumstances, for the concrete utilisation of the indicator a *norm* can be applied. This norm defines desired values and behaviour, such as “less than 10% participation of one student is too low for good collaboration”, and can be employed for providing specific messages or visualizations. In CSCL, the use of these norms is not always possible or even desirable, and thus, this element can be considered an optional aspect of the process.

13.3 Main Problems for the Integration of Interaction Analysis into Mainstream CSCL Practices

As already mentioned in the introductory section, the first objective of this chapter is to present, classify and analyze the problems that impede a wider adoption of IA tools by end users, be them researchers, practitioners or students. Taking into account the brief general overview of the interaction analysis process presented above, we present in this section a structured description of the problems met by the authors while designing and enacting CSCL scenarios. Eight of these experiences will be used to exemplify the listed problems. As it is not possible to describe in detail all the experiences that have been analyzed, we will employ one of them, from the *MosaicLearning* project (denoted as the *Mosaic* experience) to illustrate most of the problems listed in this section before we describe them with some detail in Sect. 13.3.2.

13.3.1 An Illustrating Example: The Mosaic Experience

A clear example of the problems that appear when trying to apply computer-based IA in real practice was experienced by the authors in the *MosaicLearning* research project,[1] where several groups from three Spanish universities set out to share their learning-support tools and methods (da la Fuente et al. 2008). More concretely, the overall objective was to study the issues that arise when designing, deploying and enacting a fully collaborative learning experience with remote students. The authoring

[1]http://mosaic.gast.it.uc3m.es

tool *Collage* was used to design the learning script, which was enacted by *Grail*, an IMS Learning Design run-time environment fully integrated with the .LRN Learning Management System. (.LRN 2008). Although a final general evaluation phase was also foreseen, the case study was not planned taking analysis purposes explicitly into account. This fact, together with the complexity of the final setup convert the *Mosaic* experience in a particularly good example of the many problems that can be found when researchers try to apply analysis methods to CSCL settings.

Table 13.1 describes the main aspects of this experience which is briefly introduced here. The course was an undergraduate program on Grid computing in three geographically distant higher-education institutions to a total of 12 students divided into groups of 3–4 members each (da la Fuente et al. 2008). After an initial phase of individual work, the students had to collaborate remotely in order to produce a joint conceptual map visualizing the main topics addressed by two technical reports on grid services and the service oriented computing paradigm.

In order to provide the students with a collaborative infrastructure, the team set up a VNC (*Virtual Network Computing*) server combined with single-user applications. VNC enables sharing of applications running in a remote server, which allowed us to use single user applications to support the collaborative tasks. To carry out these tasks, the students were provided with instances of *Cmaptools,* [2]

Table 13.1 Context of the *Mosaic* experience

Dimensions	Characteristics		
Scope	**Category**	**Size**	**Number of groups**
	Workgroup of 3–4 people	12 students	4 groups
Type of interaction	Distance		
Educational level	University (Graduate course)		
Tasks	**Period**	**Task**	
	Individual phase	Reading of two technical reports about the same topic and construction of a conceptual map with the main topics	
	Collaborative tasks	Build a joint conceptual map on the basis of the two previous ones	
Tools	**Tool**	**Usage**	
	Cmaptools	Construction of conceptual maps	
	Kedit	Text processing	
	Kolourpaint	Image edition	
	Skype	Remote discussions	
	Grail+.LRN	LMS engine for the deployment of the unit of learning	
Collaborative experience	Students had different previous level of experience of computer-supported collaborative learning. There was no teacher, with the participants regulating tasks, time, discussions and solutions		

[2] http://cmap.ihmc.us/

Kedit[3] and *Kolourpaint*[4] (see Table 13.1 for further details on their use). Remote communication was mediated by *Skype.*[5] It is important to note that this setup was chosen for practical reasons among other possible configurations. For example, *Cmaptools* has a collaborative version, but the design team had to discard its use, due to licensing and performance problems, forcing us to use an individual version, which was shared with the aforementioned VNC server to enable collaboration. *Kedit* and *Kolourpaint* were also a convenient choice, as they are already provided by the Linux-based environment where the experience took place.

Besides the actual design and enactment of the collaboration script, the most relevant aspect of this case for this chapter relates to the difficulties that appeared when trying to analize the experience. The team aimed to study it following the mixed evaluation method defined in Martínez et al. (2003). This method describes how to combine qualitative, quantitative and social network analysis techniques in a semi-automatic analysis based on data coming from different sources. Among these sources, the method includes data logs representing computer-mediated interactions between the participants in the experience. However, several obstacles were met when trying to collect the interactions mediated by the system in order to analyze them. First of all, .LRN only manages the use of the applications, and thus provides data on applications usage, but not about the users' actions on these applications. In fact, these actions took place through the single-user applications shared by the VNC server. However, with this VNC-based setup, the information about who originates which action was lost, which is a major obstacle to carry out any **analysis on the actors' performance**. Moreover, the actual implementation of VNC used (*tightvnc*) provides for screen capturing, which could have been combined with the audio recorded from the Skype sessions. However, it was not possible to analyze these data, because no **synchronization** mechanism between the videos from VNC and the audio from Skype had been prepared.

This experience comprises many of the problems that are normally found when researchers or educators try to apply IA to real practice. In the next subsection we describe these problems in a structured way, providing further examples taken from the authors' experience in CSCL-based projects and classroom situations.

13.3.2 A Data-Driven Analysis of Problems Regarding Integration of IA Tools and CSCL Environments

A systematic clustering of the cases reflecting authors' experience in the design and enactment of CSCL scenarios indicates that the problems found to integrate learning environments and IA tools can be classified at three levels. They may be due to

[3] http://www.kedit.com/

[4] http://www.kolourpaint.org/

[5] http://www.skype.com/

the characteristics of the applications (application level), to the actual architecture used to enable collaboration (architecture level), or to deficiencies in the overall design process (design level). This subsection elaborates on these three levels and provides further examples of cases (see also case codes in Table 13.2) where the authors met them in real practice.

Most of the cases originate from projects within the *Kaleidoscope* European Network of Excellence. In the *IA* project we worked in the integration of the teams' systems in a shared library of IA tools (Martínez et al. 2005) ([Kal-IA] in Table 13.2) where we met some of the problems discussed here. Later, in *CAViCoLA*,

Table 13.2 Problems found in integrating interaction analysis and learning environments, together with the cases in which these problems were encountered

Problem		Mosaic	Kal - CCI	Kal - IA	GS	BSCW	Wikis	Du-Pa dyads	Tagging study
Application level	Data logs are not provided or are insufficient	X				X	X		
	Data is not documented	X			X	X	X		
	Data is not processable	X							
Architecture level	Not all relevant data is captured	X	X			X	X	X	
	Data are not synchronized	X							
	Data formats are not compatible			X				X	
	The architecture does not allow to get crucial data	X							
Design level	The design did not take into account the need to analyze data	X							
	The intended object of analysis has not been integrated into the teaching practice								X
	The IA tool constraints the choice for the CSCL application					X			

we had a number of experiences to share tools and data between the participating teams. One of these experiences between Universities of Duisburg and Patras ([Du-Pa dyads] in Table 13.2) also illustrates specific problems found when working towards this integration (Harrer et al. 2006). In parallel to *CAViCoLA*, in the *CCI-IA* project ([Kal-CCI] in Table 13.2) we were asked to integrate our IA tools in the network's web-based communication and collaboration platform, in order to provide support to the members of the network. This gave us further experience in the real problems that are met when trying to apply interaction analysis tools to existing environments (Bratitsis et al. 2008). Besides these international projects, we also provide examples of local experiences where these problems or their solutions were clearly manifested, namely: the *Mosaic* experience ([Mosaic] in Table 13.2), reported in the previous subsection, the use of Group Scribbles in real classrooms ([GS] in Table 13.2), and the use of *BSCW* and *wikis* to support project- and inquiry-based learning ([BSCW] and [Wikis] in Table 13.2), and finally, an experience run in Germany with students' and teachers' tagging behavior of learning material ([Tagging study] in Table 13.2). These will be shortly described along with the problems they illustrate at the three levels that structure this subsection.

At the *application level*, it is frequent that applications **do not provide ready-to-use data** about their interactions, as was the case in the *Mosaic* experience with *Kolourpaint*, *Kedit* and the single-user version of *Cmaptools* that was employed. Another typical scenario occurs when some kind of **data** is provided, but it **does not give enough information** to perform the required analysis. For example, in the *Mosaic* experience, the data provided by *.LRN* was insufficient to perform an analysis. We have met this difficulty in several other authentic classroom experiences, where *BSCW* and *wikis* were employed to mediate collaboration. *BSCW* is a shared workspace system that provides an awareness information service to its users that can be employed for analysis. However, not all relevant data, such as repeated document readings, is provided by this service. *Wikis* usually provide ready-to-use accounts of the history of page modifications, but no data about other actions, such as readings, is recorded. Readings are usually relevant for analysis and therefore, we had to find a workaround in order to be able to analyze their data with our IA tools. In the experience based on BSCW, we could use system logs, which are not meant for end-users, but that provide useful data for analysis at a good level of detail (Martínez et al. 2003). In the *wiki-based* experience, we had to code a specific *MediaWiki* extension to be able to record readings and consider them in our analysis (Martínez-Monés et al. 2008b). Besides this, we also met difficulties to access and understand the data due to the **lack of documentation**. Thus, the use of these data for analysis requires a demanding effort and the participation of programmers before they can be employed for interaction analysis purposes. Moreover, this set of problems illustrates the fact that application builders do not think that logs are useful for end-users, and for this reason, they do not provide an easy access to them. Another example of this lack of documentation was met by the authors in their experiences with *GroupScribbles* ([GS] in Table 13.1), a tool that enables students and a teacher to share contributions on sheets similar to post-it notes and to jointly manage the movement of these electronic notes within and between public and private spaces

(Roschelle et al. 2007), (Dimitriadis et al. 2007). The first version of this tool, based on the *TupleSpaces* architecture, provided usage logs only for internal technical debugging purposes, and therefore the lack of an adequate documentation impeded the use of these logs for analysis. The net result of this situation is that valuable information on the interactions among students and teachers could not be used for analysis. The last problematic issue that can be found at the application level is that the **data** itself might **not be directly processable** by computer tools in order to extract indicators. This is the case with streamed data such as audio, video, etc., like the data provided by *Skype* and *VNC* ([Mosaic]). These data is appropriate for a thorough review of the experience, but it is not directly understandable by an automatic data analysis application. It requires human intervention, in the form of coding or labeling before this data is usable for computing IA indicators. In fact, this is an approach followed by many IA tools aimed to supporting research, like *ActivityLens* (Fiotakis et al. 2007) or Tatiana (Dyke 2008) but it becomes an open issue how these data is meant to be prepared and handled by practitioners, which do not have the time and resources needed to process them.

Some of these problems at the application level are translated to the *architecture level*, i.e., to the actual configuration of teaching/learning tools set up in order to support the collaborative tasks. For example, even if the experience uses an application with a complete and well documented data log, it might happen that this application covers only part of the interactions, and therefore, the **data provides only a partial view** of the collaborative tasks carried out by the students. In the Mosaic experience, we could not use *.LRN* to store and analyze the participants' interactions because this system was used to launch applications, but not to mediate interactions. Examples of this problem can be found very frequently, as students' communication is normally not controlled, and happens outside the system many times. This happened in the aforementioned experiences, where we employed *BSCW* and *wikis* to support the collaborative tasks. In spite of the mentioned difficulties, we were able to analyze interactions mediated by these platforms, but the most of the students' internal communication was mediated by e-mail or instant messaging, which was not possible to record, and was thus not included in the analysis. Another case where this problem was met was in several cross-site studies between universities of Duisburg and Patras ([Du-Pa dyads] in Table 13.2) (Harrer et al. 2006), where the remote interaction, especially the use of their external chat tools, between students was not always possible to register, and thus, it could not be analyzed. A third example of this was experienced by the authors in the CCI-IA project, where we were asked to provide an IA service for the Kaleidoscope site ([Kal-CCI] in Table 13.2), based on the analysis of the interactions on the platform. Again, the main problem found in this project was that most of the interactions among teams happened outside the system (Bratitsis et al. 2008).

A second problem to consider at the architecture level happens when the data is stored, but it is not possible to integrate them because they are not **semantically** compatible. A very clear example of this issue was met by the authors in the aforementioned effort to build a library of shared IA tools ([Kal-IA] in Table 13.2). This effort was hindered by the fact that their data models were not compatible. A more concrete example of this problem was met in the mentioned cross-site studies between Duisburg and Patras, where it was difficult to integrate process data of the students' interaction with their products, etc. Related to this integration problem we can find the difficulties to **synchronize** data from different sources. Synchronization is feasible provided some previous requirements are met, such as that the data is time-stamped with a common time reference. Then, there exist software packages specifically oriented to synchronize these data and show it to the researchers for its coding and analysis. However, as shown by the *Mosaic* experience, this data-stamping requires an extra effort to the designers which is not always possible. Finally, it might happen that the **overall architecture** used to enable collaboration **hinders analysis**. This was clearly reflected in the *Mosaic* experience previously described. The VNC-based implementation used did not allow distinguishing who executed the actions, which is a major problem for the later analysis of these data.

A major source of problems related to the integration of IA in real practice deals with the *design level*. First, when setting up a scenario to enable and study collaboration, the aspects related to interaction analysis are not normally considered a first priority, and many times this means that they are neglected in benefit of more immediate requirements, such as the need to provide for collaboration-enabling functionalities and an acceptable performance. This was the main reason why the Mosaic setup did not facilitate the recording of interactions, and there are many other examples where this lack of priority ends up with a setup that does not allow researchers or educators to study the case in its fully extent. This is a very common situation in every innovative project, but it would be a minor problem if the tools provided ready-to-use logs. Therefore, this shows that the problems at the tool level are also reflected at this one. Another issue that must be considered at design level is whether or not the actual analysis objectives are possible to reflect in the teaching practices. The need of using real scenarios to analyze CSCL practices often meets the obstacle that it is **difficult to integrate these aspects in normal teaching practices** and environments, which tend to be slow in innovation. An example of this problem was found by the authors in an experience ([Tagging study] in Table 13.2), where the phenomenon of social tagging of learning material was explored (Lohmann et al. 2008): since the e-learning platform used at that university did not integrate a tagging feature, the experiment was emulated outside of the learning platform in a pen-and-paper pre-study. Even when this integration has been achieved and an authentic scenario is found where teaching practice meets research objectives we can find a new problem. As illustrated by the *ICALTS* model described in Sect. 13.2, the learning environment can be restricted to those that provide the data needed by the analysis, thus restricting the choice of the collaboration-supporting tools to those that might be not the most appropriate ones for the planned tasks. An example of this case was met in the *BSCW*-based experience, where we could not upgrade the system to a newer version, which did no longer provide the system logs we were using for analysis. This problem has its roots on the fact that not all collaboration-support tools provide ready-to-use log files, or that these logs are not compatible among each other.

The above data-driven analysis provides an overview of the problems encountered, when researchers or educators want to integrate interaction analysis techniques and tools in CSCL environments. The conceptual assignment of problems to three levels aims at analyzing the common issues and providing a global explanation of their origin. However, the above analysis provides sufficient hints on the existing correlation among levels, as well as the actors involved. In this sense, the design level originates many of the problems related to the educators or researchers, i.e., the practitioners or end-users, while the application and architectural levels deal with computer scientists and engineers, i.e., the technology providers of CSCL environments and IA tools. Apparently, there is a significant gap between the two worlds which impedes the solution of the observed problems.

The following section points to some solutions that may bridge the gap and allow for a seamless integration, taking into account trends and advances in different fields. These may focus either at the design level (integrated or multi-perspective scripting or learning design), application (IA-aware tools and learning environments) or at the architectural level (decoupled and service-oriented architectures and common data protocols).

13.4 Towards an Integrated Perspective on Learning and Analysis Activities in CSCL

From the problems described in the previous section we can derive a need to address the issue of interaction analysis early in the preparation of the learning activities. Our first ideas on that have been described in Martínez-Monés et al. (2008a) and will be expanded in the following. As a first differentiation we propose solutions that are mainly based on direct consideration during the *design* process and solutions that are mainly based on preparations of the learning *technology*, i.e., the tools used. The design-driven solutions already try to take into account the design level issues of the previous section, i.e. they explicitly make the need for analytical activities visible in the design process. The technology-driven solutions mainly tackle the application and architecture level by means of providing well-defined interchange formats between learning and analysis tools.

13.4.1 Design-Driven Solutions

For the first type of proposals the needs of interaction analysis are directly integrated into the design process, which means that the activities related to the analysis (like assessment, evaluation, research or monitoring) are explicitly modeled and specified, together with the rest of the learning process before the activity is implemented. Otherwise the problems raised in the previous section might arise, when studying the collected data, possibly finding shortcomings in the richness, availability etc.

Since the concerns of the learning process and those of analysis can have different needs and issues, multiple and potentially conflicting aims have to be addressed. A usual means to tackle this integration problem in computer science is either to follow a *co-design* approach that directly integrates the diverse needs or to use a *multi-perspective* approach where each facet is represented individually and relations between the different perspectives are made explicit by meta-rules or constraints. Prominent examples for these different approaches are hardware/software co-design or the multi-perspective software modeling language UML (unified modeling language with multi-perspective diagrams). Interestingly, this differentiation has also been discussed in the area of learning design (Botturi and Stubbs 2007) where the dimension "perspective" has been used to differentiate between single-perspective and multiple-perspective approaches to represent learning designs. For the integration of the needs that interaction analysis brings forth we use a similar differentiation.

In a co-design approach of the learning and analysis processes, the aspects of the *learning/teaching* activities and the *observation/analysis* are taken into account simultaneously, posibly by a single designer. One design process using this approach is described in Villasclaras et al. (2009) where both aspects are represented by means of a pattern language: collaborative learning flow patterns (CLFP) and assessment patterns describe at an abstract level the essential characteristics of the activities. While a CLFP describes the learning/teaching activities, the assessment patterns describe the activities related to the collection of information for assessing the students, i.e., an observation/analysis task. According to this approach, the designer of a learning scenario chooses and configures both types of patterns in an integrated process. Since there might be constraints between specific learning and assessment activities, these constraints have to be represented in a pattern language to inform the designer of potential problems when using patterns of different type in combination. Figure 13.2 shows an example of how learning

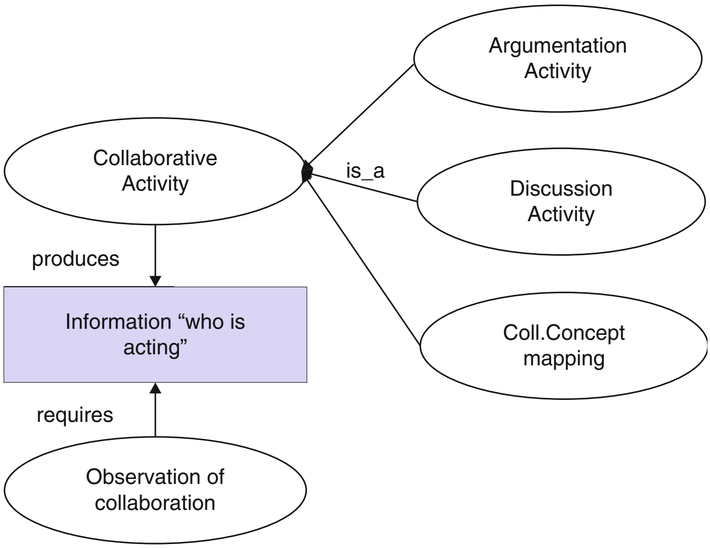

Fig. 13.2 An example of joint study and design of learning and interaction analysis activities

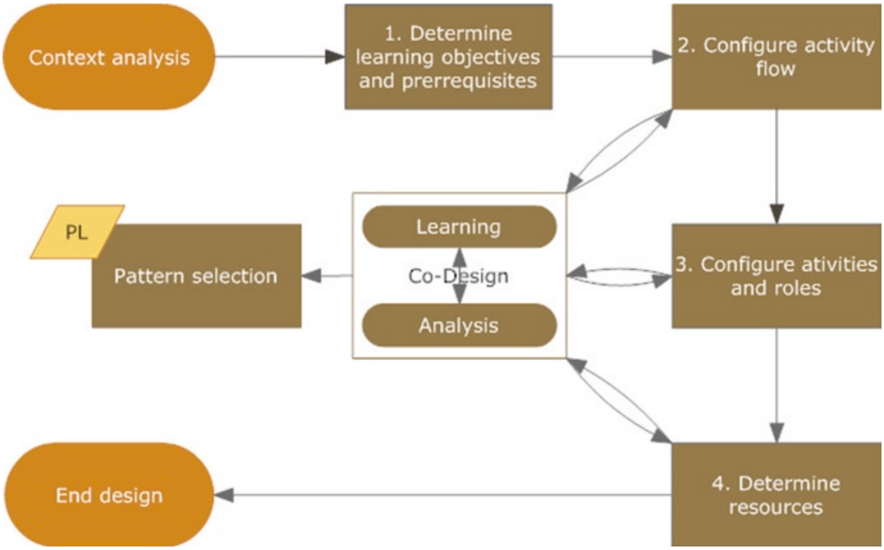

Fig. 13.3 A unified view of the co-design process (Adapted from Villasclaras et al. (2009))

activities (at the top of the figure) and analysis activities can be related to each other by high-level dependencies helpful for coordinated planning of both aspects: the choice of a collaborative concept mapping creates a potential information “who is acting” that can be used by an observation activity.

Thus, it is reasonable to think of following a co-design process such as the one shown in Fig. 13.3, in which learning and analysis needs are taken into account in an integrated way. An example of a concrete solution that could be produced with this co-design process is shown in Fig. 13.4, where the peer-review assessment pattern (that can be seen as a specific kind of interaction analysis activity performed by a peer student) is designed to be used together with the pyramid, jigsaw and think-pair-share patterns, which are specific cases of CLFP, i.e., learning patterns. For more information on these pattern-based solution, see Villasclaras et al. (2009).

An advantage of this approach is that the information given in the relations between learning activities and analysis activities helps to get a harmonized and coherent process model immediately, because inconsistencies would be visible to the designer automatically. Drawbacks of this co-design approach are on the one hand that the expertise in both fields is needed at once, so a division of labor is hard to achieve, and on the other hand, that the methodology of design is “closed” in that respect, that other approaches for learning design or analysis cannot be combined with it.

In the multi-perspective approach the learning process and the analysis process are modeled separately with a method of the designers’ choice, where each aspect can be designed by a different expert of the respective field. This potential of using division of labor between experts in the different aspects and also allowing each expert to use a method of her/his choice is a substantial advantage of this approach. The main challenge with this approach is the integration of the learning and

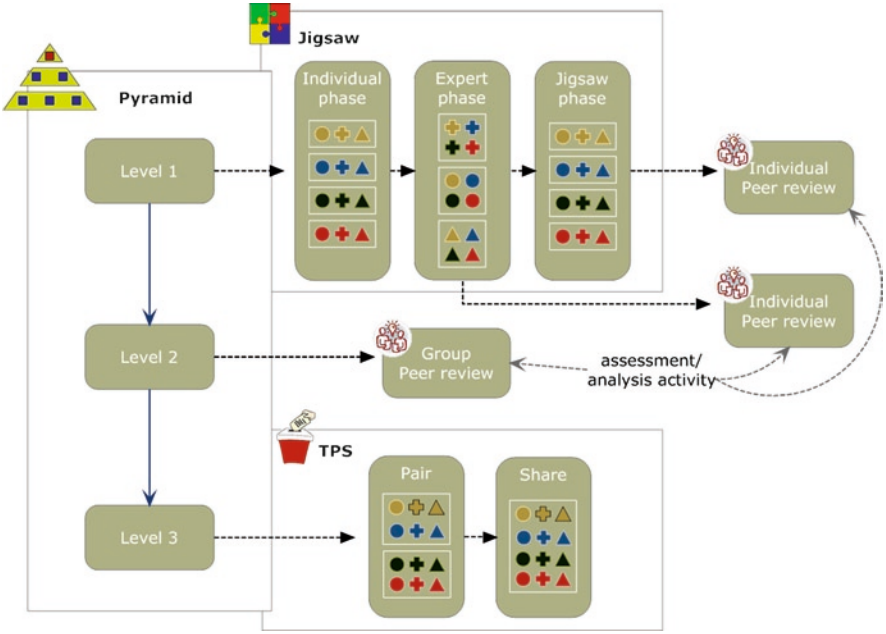

Fig. 13.4 An example of co-design that takes into account learning and assessment activities

analysis processes into one model that takes into account both perspectives appropriately. For this end, high-level constraints between the two perspectives are needed to allow a meaningful integration of these. This gap between the two different perspectives can be bridged by an abstraction level for learning tools and analysis tools, so that no concrete tools are defined in the respective processes: when both the learning activities and the analysis activities are modeled with their abstract purpose/goals, a suitable combination of learning and analysis tools can be searched for using a categorical representation of tools. One representative of this approach is the *OntoolCole/Ontoolsearch* (Vega-Gorgojo et al. 2008) environment, where an ontology of learning tools helps to categorize specific tools according to their general purpose. The extension with a similar categorization schema for analysis tools would help to inform the designers of a multi-perspective modeling approach if their modeled design can be conducted with the available selection of learning and analysis tools and what a recommended combination could be.

As a reflection on the *Mosaic* example from the previous Sect. 13.3.1, of the problems for analysis was the lack of logfile information with respect to the actor of an action. Using the *OntoolCole/Ontoolsearch* approach a designer would specify “collaborative concept mapping tool” and “tool for observation of collaboration” as required for the experience. Table 13.3 gives a simple list for several tools available for the activity that could be provided by *OntoolSearch*:

Based on the categories assigned to available tools, a possible recommendation would be the collaboration tool *FreeStyler* with its concept mapping functionality

Table 13.3 Selection of appropriate tools for the Mosaic experience based on categorization of learning and analysis tools

Tool	Category	Format produced	Format consumed
Cmaptools + VNC	Collaborative concept mapping	VNC log [without user info]	
FreeStyler + concept map plug-in	Collaborative concept mapping	Common format logs [with user info]	
.LRN monitoring	Observation of student progress		IMS Learning Design
Argunaut system	Observation of collaborative activities		Common format logs

and logfile information in the *common format* defined in the Kaleidoscope IA project, and the *Argunaut* observation tool that consumes logfiles in this *common format*. The combination of the tools would be a recommendation, because the tools not only comply to the specified categories (this would also fit for the combination of *Cmaptools/VNC* and *.LRN*) but also with respect to the logfile format, which facilitates the integration of both into an integrated learning/analysis process (which was not possible in the *Mosaic* experience with the used tools).

Besides this challenge how to integrate the two different perspectives, the latter multi-perspective approach has the benefit that each perspective can be modeled separately by an expert in that field using the design method of her/his choice, i.e. the flexible combination of different methods for the learning design and the analysis process is possible, which would also make comparative analyses feasible, e.g. using two different analyses processes with the same learning design method or vice versa.

13.4.2 Technology-Driven Solutions

Technological infrastructure can be a show-stopper or enabler of analysis processes in computer-based learning scenarios. As motivated in the previous section, the analysis process can be severely compromised by the technology used, if important data is missing (such as in the *Mosaic* example), heterogeneous data cannot be synchronized and/or integrated etc.

Yet, there are several provisions that can be made to prepare the infrastructure for a support of both learning and analysis processes: **interoperability** is the main principle that allows conducting learning and analysis processes in a coordinated manner. Besides some systems that have tightly-coupled learning tools and analysis tools (e.g. the *Synergo* system (Avouris et al. 2004)) that has a built-in analysis tool for the teacher/researcher, usually the learning tools and the analysis tools are not within the same codebase and developed by the same teams. Thus, data exchange and semantic interoperability between the different tools are a pre-requisite to conduct the analysis process.

The idea of having independence between the CSCL and IA-related codebases leads to **decoupled architectures**, where both systems are able to run independently of each other, but where the semantics and syntax of the interactions are shared, so that the analysis processes can get the most of the solutions. We can distinguish two approaches. The first is based on inter-process communication, while the second is based on interchange of log file information. Figure 13.5 shows the schema for both types of decoupled architectures. Examples for the first approach have been proposed for educational systems that combine the functionality of different stand-alone applications. In Ritter and Koedinger (1997) the combination of simple learning tools with the diagnostic capabilities and feedback messages of Cognitive Tutors has been discussed, which opens up the possibility of on-the-fly analysis and tutoring feedback for the integrated system. Service-oriented approaches are well-suited for this type of integration, because a (web) service provides a well-defined interface description for the data exchange and its granularity is well-suited to compose complex educational applications from several services. The use of (web) services for educational systems has been discussed early in Chen (2003) and Vaquero-González et al. (2005) and has been followed up by several implementations of educational systems, such as GridCole (Bote-Lorenzo et al. 2008) and Finesse (Allison et al. 2005). Currently, these systems still have a limited scope, because the number of existing learning services and especially of analysis services is too low to allow a variable mix-and-match of services for flexible construction of learning scenarios. Given the expected larger set of educational services, the instantiation of services for learning scenarios via graphical editors is a promising approach to relieve the scenario designers of deep technical knowledge. A similar approach to this can be found in existing learning environments, like Moodle, that enables its users (mostly teachers) to instantiate specific tools, such as chats, wikis, etc.

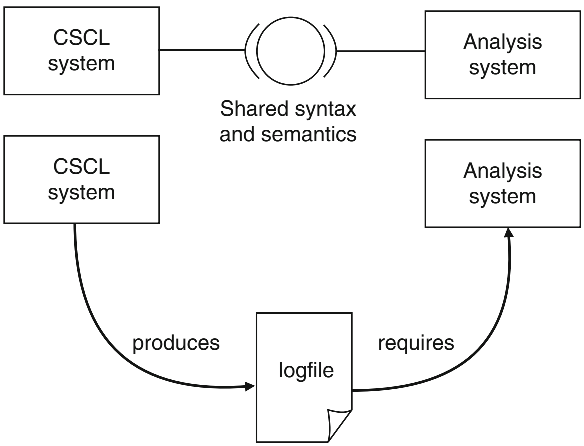

Fig. 13.5 Schema of the two decoupled architectures. On *top* of the diagram the inter-process communication via communication of shared syntax and semantics. At the *bottom* the coupling via logfiles

Log files can also be the base of decoupled architectures, provided that the **semantics** of the log events are known and shared between the CSCL and the IA sub-systems. This would be a step forward in the use of log files, so that different IA tools would be able to be used with different CSCL systems and vice versa, providing for a flexible mix-and-match. To facilitate the flexible combination of different analysis tools during the process, the international initiatives ICALTS, IA, and CAViCoLA between several European research teams defined the standardized common data format that captures the relevant information of collaborative learning activities for follow-up analyses. A detailed description of this format is out of the scope of this chapter, but the interested reader can find it in Harrer et al. (2009). The *Argunaut* (De Groot et al. 2007) system is an example where the standardization of log files has been used as the mediating vehicle to allow moderators (e.g., teachers) of synchronous discussions the monitoring and evaluation of the ongoing discussion(s). Different discussion environments can be integrated into the system, if they provide the log file format as output (Harrer et al. 2008b). The technical framework of the *Argunaut* system is shown in Fig. 13.6, where arbitrary discussion environments can be integrated into the system, given that they comply with the defined logfile format (used by the Protocol Processor) and implement the desired

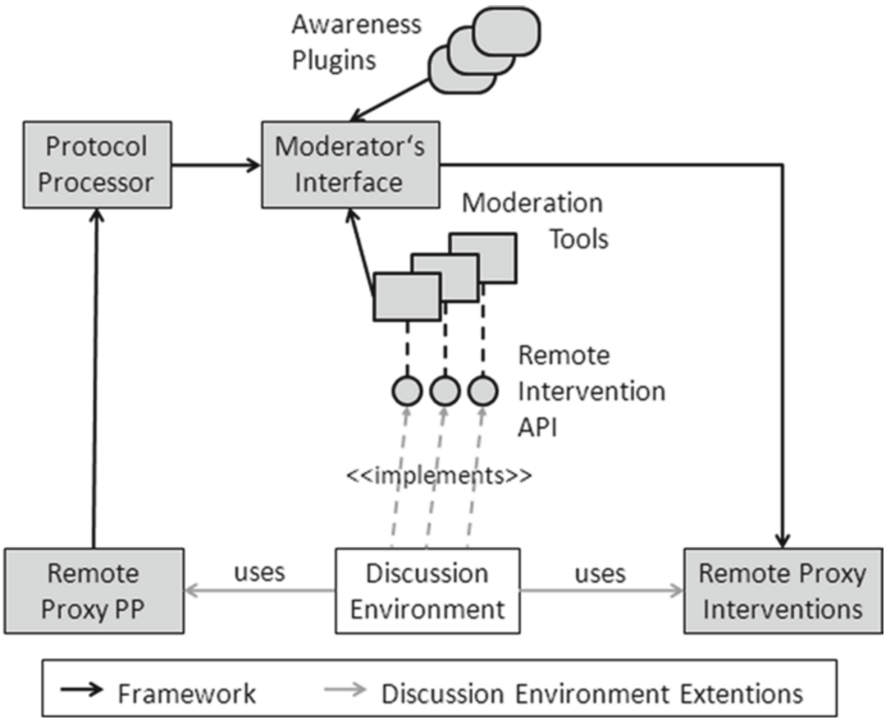

Fig. 13.6 Argunaut framework, showing how different discussion environments can be plugged into the environment

moderation features (Remote Intervention API) to allow the moderator of e-discussions the intervention into ongoing discussions. Indeed the framework provides a solution that uses inter-process communication via the proxies to allow on-the-fly observation and intervention of discussions based on the exchange of common format events in the well-defined logfile format.

A generic processing scheme for analysis has been enabled due to this common data format: Its validity has been tested through several CSCL and IA tools within the Kaleidoscope network (Kaleidoscope 2007), including heterogeneous indicators based on Social Network Analysis (SNA) or qualitative methods. Other initiatives in this direction are the MULCE project, that aims to define a learning data corpora for sharing purposes (Chanier et al. 2009), and the *Centralised Research Data Repository*, another Kaleidoscope initiative, which aimed to define a common ontology to share learning materials among researchers (Centralized Research Data Repository 2007).

Because in some cases the modification of the original logfile formats is not desirable or brings a substantial effort with it, the use of adapter components is a potential technical solution for this problem: with the help of adapters the data sources/logs can be made compatible with each other so that analysis tools can be used with learning tools of a third party. The practical usage of adapters in heterogeneous educational scenarios (i.e. using several independent learning tools) has been demonstrated and discussed recently in Harrer et al. (2008a). The effort needed to generate an adapter component is relatively low compared to refactoring tools into full-fledged web or grid-services, which makes this proposal a good alternative for initial rapid development. The use of a mediator was also proposed by the aforementioned *centralized data repository* initiative, in order to enable tools with different underlying ontologies to access a **common ontology** representing a wide range of possible learning objects.

In spite of their partial success, these proposals still represent local efforts, and show that there is a need for an agreement by the various stakeholders of the community that would allow for a generalized sharing of tools and data among teams.

13.5 Conclusions

Computer-supported interaction analysis tools and methods have the potential to leverage research and practice in CSCL. This fact has raised the interest of the research community, which has been reflected in a growing number of research projects, meetings and papers focused on these themes. However, current practice is not benefiting from the potential advantages of applying IA tools to their settings. Among the reasons that explain this mismatch we can identify problems at the design level, where end users such as researchers or practitioners do not plan in advance for configurations that allow for interaction analysis. Sometimes, these problems are related to issues at the application and architectural levels, where (learning) system developers do not provide for ready-to-use interaction data or if they do, they do not worry about their interoperability or synchronization with other sources of data.

Several lines of work can help to overcome these problems. At the design level, interaction analysis issues must be integrated in the overall design process. This can be done following a co-design approach or a multiple perspectives approach, taking into account the trade-off between a consistency control and a division of labor at design time. At the application and architecture levels, technology-driven solutions based on decoupled architectures are feasible. These architectures can be implemented following an inter-process communication or a log-file interchange approach, aiming at enhancing interoperability while fostering integrated use of CSCL and IA tools.

The problems stated in this chapter reflect a large variety of situations in both European and national projects, and heterogeneous or homogeneous design teams. These problems are expected to be even worse in the case of a wider adoption of the IA tools and techniques by practitioners. Then, this review aims to raise the awareness of all the implied actors on the issues that must be taken into account to increase the use of IA tools in real CSCL settings. This is a noteworthy effort; as it would allow researchers and practitioners improve their experiences by being able to reflect on them and by adding new monitoring and assessing capabilities to their CSCL settings.

However, several issues have been detected that remain unsolved, and call for further efforts in the area. First of all, we need to increase interoperability among our data, but the complexity and richness of CSCL settings make it difficult to reach a common agreement on these data. This interoperability might be achieved following different paths, from minimalist approaches that define the minimum set of data needed by tools to inter-operate, or by means of practical approaches that propose the use of adapters to match local formats (enabling for specific types of analysis) to a generic one (enabling for sharing data and tools). Of course, these are not easy approaches and many issues remain unsolved that need further discussion and agreements among the community which has been started in recent years with several international workshops and initiatives on analysis methods and integration of methods with different tools (Dwyer et al. 2008; Law et al. 2009)

Improvements at the technological level are also needed. Especially, there is a need to increase the number of IA tools and services that could then be chosen by researchers and practitioners in the interaction-aware design processes outlined in this chapter. A major adoption of these design processes could also be benefited by tools, such as the one proposed by Villasclaras et al. (2009), for integrating assessment into the learning design processes.

Acknowledgments The authors would like to thank their colleagues for their intensive cooperation carrying out the cases mentioned in this chapter. This work has been partially supported by the Spanish Ministry of Science and Education (project TIN2008-03023/TSI) and the autonomous government of Castilla and León, Spain (VA106A08, VA107A08).

References

LRN (2008). http://dotlrn.org/ Accessed 23 Aug 2009.

Allison, C., Nicoll, J. R., Bateman, M., & Ling, B. (2005). From web to grid: Evolution of a learning environment. In *The 4th IASTED International Conference on Web Based Education WBE'05. 2005*. Grindelwald: ACTA.

Avouris, N., Margaritis, M., & Komis, V. (2004). Modelling interaction during small-group synchronous problem-solving activities: The Synergo approach. In *2nd International workshop on designing computational models of collaborative learning interaction at ITS 2004, 7th Conference on Intelligent Tutoring Systems*. Maceio, Brazil.

Bote-Lorenzo, M. L., Gómez-Sánchez, E., Vega-Gorgojo, G., Dimitriadis, Y., Asensio-Pérez, J. I., & Jorrín-Abellán, I. M. (2008). Gridcole: A tailorable grid service based system that supports scripted collaborative learning. *Computers & Education, 51*(1), 155–172.

Botturi, L., & Stubbs, S. T. (Eds.). (2007). *Handbook of visual languages for instructional design: Theories and practices*. Hershey: Idea Group.

Bratitsis, T., Dimitracopoulou, A., Martínez-Monés, A., Marcos-García, J. A., & Dimitriadis, Y. (2008). Supporting members of a learning community using interaction analysis tools: The example of the Kaleidoscope NoE scientific network. In *8th IEEE International Conference on Advanced Learning Technologies (ICALT 2008)*. Santander: IEEE Computer Society.

Centralized Research Data Repository (2007). http://www.noe-kaleidoscope.org/group/datarep/ Accessed 23 Aug 2009.

Chanier, T., Reffay, C., Lamy, M. N., & Ganascia, J. G. (2009). Project MULCE (Multimodal Learning Corpus Exchange) http://mulce.univ-fcomte.fr/. Accessed 23 Aug 2009.

Chen, W. (2003). Web services in web-based education systems. *In Proceedings of Computers and Advanced Technology in Education, CATE/WBE '2003*. Rhodes: Greece.

De Groot, R., et al. (2007). Computer supported moderation of e-discussions: The ARGUNAUT approach. In C. Chinn, G. Erkens, & S. Puntambekar (Eds.), *Computer-Supported Collaborative Learning: Mice, Minds and Society. Proceedings of the Seventh International Computer Supported Collaborative Learning Conference (CSCL 2007)*. Mahwah: Lawrence Erlbaum.

Dimitracopoulou, A. (2005). Designing collaborative environments: Current trends and future research agenda. In T. Koschmann, T.-W. Chan, & D. D. Suthers (Eds.), *Computer Supported Collaborative Learning 2005: The next 10 years!* Taipei: Lawrence Erlbaum.

Dimitriadis, Y., Asensio-Pérez, J. I., Hernández-Leo, D., Roschelle, J., Brecht, J., & Tatar, D. (2007). From socially-mediated to technology-mediated coordination: A study of design tensions using Group Scribbles. In C. Chinn, G. Erkens, & S. Puntambekar (Eds.), *Computer-Supported Collaborative Learning: Mice, Minds and Society. Proceedings of the Seventh International Computer Supported Collaborative Learning Conference (CSCL 2007)*. Mahwah: Lawrence Erlbaum.

Dyke, G. (2008). Tatiana: Trace Analysis Tool for Interaction ANAlysts, http://tatiana.emse.fr/Presentation/summary.html. Accessed 23 Aug 2009.

Dwyer, N., Law, N., Rose, C.P., & Suthers, D. (2008). *Workshop: A common framework for CSCL interaction analysis at International Conference of the Learning Sciences, ICLS 2008*, Utrecht, The Netherlands. http://engaged.hnlc.org/stories. Accessed 23 Sep 2009.

Fiotakis, G., Fidas, C., & Avouris, N. (2007). Comparative usability evaluation of web systems through ActivityLens. In *Proceedings of the 11th Panhellenic Conference on Informatics (PCI 2007)*. Patras, Greece.

da la Fuente, L., Pardo, A., Delgado, C., & Asensio, J. I. (2008). Collaborative learning models on distance scenarios: A case study. *In 8th IEEE International Conference on Advanced Learning Technologies, (ICALT 2008)*. Santander: IEEE Computer Society.

Guribye, F., & Wasson, B. (2002). The ethnography of distributed collaborative learning. Paper presented at the Computer supported collaborative learning 2002.

Harrer, A., Kahrimanis, G., Zeini, S., Bollen, L., & Avouris, N. (2006). Is there a way to e-Bologna? Cross-national collaborative activities in university courses. *In Proceedings of the 1st European Conference on Technology Enhanced Learning (EC-TEL)*. Crete, Greece.

Harrer, A., Zeini, S., Kahrimanis, G., Avouris, N., Marcos, J. A., Martínez-Mones, A., et al. (2007). Towards a flexible model for computer-based analysis and visualization of collaborative learning activities. In C. Chinn, G. Erkens, & S. Puntambekar (Eds.), *Computer-Supported Collaborative Learning: Mice, Minds and Society. Proceedings of the Seventh International Computer Supported Collaborative Learning Conference (CSCL 2007)*. Mahwah: Lawrence Erlbaum.

Harrer, A., Pinkwart, N., McLaren, B. M., & Scheuer, O. (2008). The scalable adapter design pattern: enabling interoperability between educational software tools. *IEEE Transactions on Learning Technologies, 1*(2), 131–143.

Harrer, A., Ziebarth, S., Giemza, A., & Hoppe, U. (2008). A framework to support monitoring and moderation of e-discussions with heterogeneous discussion tools. In *Proceedings of the 8th IEEE International Conference on Advanced Learning Technologies ICALT 2008*. Santander: IEEE Computer Society.

Harrer, A., Martínez-Monés, A., & Dimitracopoulou, A. (2009). Users' data: Collaborative and social analysis. In N. Balacheff, S. Ludvigsen, T. de Jong, A. Lazonder, & S. Barnes (Eds.), *Technology-enhanced learning: Principles and products*. Berlin: Springer.

Kaleidoscope (2007). Kaleidoscope NoE. Projects: ICALTS (Interaction & Collaboration Analysis' supporting Teachers & Students' Self-regulation) and IA (Interaction Analysis' supporting teachers & students' self-regulation), and the European research team CAViCoLA (Computer-based Analysis and Visualisation of Collaborative Learning Activities), from http://www.noekaleidoscope.org/. Accessed 23 Aug 2009.

Law, N., Suthers, D., Lund, K., Rosé, C., & Teplovs, C. (2009). Common Objects for Productive Multivocality in Analysis. Workshop at the Computer Supported Collaborative Learning Conference 2009. Rhodes, Greece.

Lohmann, S., Thalmann, S., Harrer, A., & Maier, R. (2008). Learner-driven annotation of learning resources - lessons from experiments on tagging. In *Proceedings of the 8th International Conference on Knowledge Management I-KNOW 08*. Graz, Austria.

Markauskaite, L., & Reimann, P. (2008). Enhancing and scaling-up design-based research: The potential for E-Research. In *Proceedings of the International Conference of Learning Sciences, ICLS 2008*. Utrecht.

Martínez, A., Dimitriadis, Y., Rubia, B., Gómez, E., & de la Fuente, P. (2003). Combining qualitative evaluation and social network analysis for the study of classroom social interactions. *Computers & Education, 41*(4), 353–368.

Martínez, A., Harrer, A., Barros, B., & Vélez, J. (2005). Library of Interaction Analysis Methods Kaleidoscope NoE.

Martínez-Monés, A., Dimitriadis, Y., & Harrer, A. (2008). Interaction-aware design for learning applications. Reflections from the CSCL field. In *Proceedings of the 8th IEEE International Conference on Advanced Learning Technologies, ICALT 2008*. Santander: IEEE Computer Society.

Martínez-Monés, A., Villagrá-Sobrino, S., Santos-Fernández, R., Anguita-Martínez, R., & Jorrín-Abellán, I. M. (2008b). Social network analysis support for an IBL wiki-based course. Workshop: Real-time methods, at International Conference of the Learning Sciencies, ICLS 2008, Utrecht.

Moreno, C., & Ventura, S. (2007). Educational data mining: A survey from 1995 to 2005. *Expert Systems with Applications, 33*, 135–146.

Ritter, S., & Koedinger, K. (1997). An architecture for plug-in tutor agents. *International Journal of Artificial Intelligence in Education, 7*(3/4), 315–347.

Roschelle, J., Tatar, D., Chaudhury, S. R., Dimitriadis, Y., Patton, C., & DiGiano, C. (2007). Ink, improvisation, and interactive engagement: Learning with tablets. *Computer, 40*(9), 38–44.

Soller, A., Martínez, A., Jermann, P., & Muehlenbrock, M. (2005). From mirroring to guiding: A review of the state of the art technology for supporting collaborative learning. *International Journal of Artificial Intelligence in Education, 15*, 261–290.

Vaquero-González, L. M., Hernández-Leo, D., Simmross-Wattenberg, F., Bote-Lorenzo, M. L., Asensio-Pérez, J. I., Dimitriadis, Y., et al. (2005). The opportunity of grid services for CSCL-application development. In *Proceedings of the 13th Euromicro Conference on Parallel, Distributed and Network-Based Processing, PDP'05*. Lugano.

Vega-Gorgojo, G., Bote-Lorenzo, M. L., Gómez-Sánchez, E., Asensio-Pérez, J. I., Dimitriadis, Y., & Jorrín-Abellán, I. M. (2008). Ontoolcole: Supporting educators in the semantic search of CSCL tools. *Journal of Universal Computer Science, 14*(1), 27–58.

Villasclaras, E., Hernández, D., Asensio, J. I., & Dimitriadis, Y. A. (2009). Incorporating assessment in a pattern-based design process for CSCL scripts. *Computers in Human Behavior, 25*(5), 1028–1039.

Chapter 14
A Framework for Assessment of Student Project Groups On-Line and Off-Line

Gahgene Gweon, Soojin Jun, Joonhwan Lee, Susan Finger, and Carolyn Penstein Rosé

Abstract Assessment of difficulties within group processes, especially through automatic means, is a problem of great interest to the broader CSCL community. Group difficulties can be revealed through interaction processes that occur during group work. Whether these patterns are encoded in speech recorded from face-to-face interactions or in text from on-line interactions, the language communication that flows between group members is an important key to understanding how better to support group functions and therefore be in a better position to design effective group learning environments. With the capability of monitoring and then influencing group processes when problems are detected, it is possible to intervene in order to facilitate the accomplishment of a higher quality product. In this chapter we address this research problem of monitoring group work processes in a context where project course instructors are making assessments of student group work. Thus, our purpose is to support those instructors in their task. We describe the mixed methods approach that we took, which combines both an interview study and a classroom study. Three research questions are answered: (1) What do instructors want to know about their student groups? (2) Is the desired information observable, and can it be reliably tracked by human annotators? (3) Can the desired information be automatically tracked using machine learning techniques to produce a summary report that instructors can use? Based on interviews with nine instructors, we identified five process assessment categories with subcategories at the group and individual level: namely, goal setting, group and individual progress, knowledge contribution, participation, and teamwork. We verified that these assessment categories can be reliably coded during group meetings with a reliability of $r=0.80$ at the group level and $r=0.64$ at the individual level using carefully constructed human assessment instruments. We present work in progress towards automation of this assessment framework.

G. Gweon (✉)
Human–Computer Interaction Institute, Carnegie Mellon University,
5000 Forbes Ave., Pittsburgh, PA 15213-3891, USA
e-mail: gkg@cmu.edu

S. Puntambekar et al. (eds.), *Analyzing Interactions in CSCL*, Computer-Supported Collaborative Learning Series, DOI 10.1007/978-1-4419-7710-6_14,

14.1 Introduction

Whether face-to-face or interacting on-line, group collaboration is difficult, and effective monitoring and support is necessary in order to ensure success as much as possible. Thus, the problem of automatic assessment of group work processes from recorded data, whether in the form of speech recordings, or in the form of text based interactions on-line, is of great interest to the CSCL community. In this chapter we describe our work on automatic assessment of group processes from face-to-face interactions recorded as digitized speech. However, we consider this work to be a model that could be followed by CSCL researchers to study groups in other contexts and apply machine learning technology to the specific problems of real time assessment of group processes that are present in the contexts in which they are doing their work. We present a three step methodology in which an interview study informs the development of the target assessment criteria, a classroom study provides data for developing instruments for hand annotation of data, and further data collection provides the data with which to build an annotated corpus. Finally, machine learning technology is applied to automatically replicate the assessments made by hand, which conform to the criteria obtained in the interview study. We offer this chapter as a road map to CSCL researchers who are interested in embarking upon a similar journey.

Group difficulties can be revealed through interaction processes, which display aspects of group dynamics, such as amount of effort offered by group members or level of participation. If it were possible to trace these processes, group overseers such as managers or instructors, would be in a better position to influence these processes to a positive end. It would then be possible to enhance both the learning experience for group members as well as to facilitate the accomplishment of a higher quality product. Given the importance of maintaining effective group processes, there has been much recent interest analyzing and monitoring traces of group activities that are predictive of group process assessments (Madan et al. 2004; Rienks et al. 2006). Researchers have traditionally used qualitative and quantitative methods to assess group processes by manually coding for these types of events (Meier et al. 2007; Weinberger and Fischer 2005). More recently, others have used machine learning and data mining technology to analyze and monitor group processes. The chapter by Reimann, Yacef, and Kay in this book also uses data mining methods for analyzing changes that occur in groups. However, other automatic assessment work has tended to use data collected in limited environments, either in a laboratory setting (Rienks et al. 2006) or in a small number of single class sessions (Chen 2003). Work focused in this way serves as a proof of concept of the technology and a good starting point for a longer term investigation, since it targets short-term goals that are attainable with current technology. However, in both the manual coding and data mining approaches, evidence that this early work addresses the real issues faced by project course instructors in their assessment work is lacking. In particular, it is unclear whether the group processes addressed by researchers correspond to the most important ones that instructors would use in a classroom environment.

In this chapter, we present a mixed methods approach used to develop a computer-based assessment framework for automatic, unobtrusive monitoring of group interactions in face-to-face group meetings. The purpose is to enable project course instructors to intervene in a more timely manner to support effective group processes. Our approach differs from the existing literature in that we use an interview study to motivate the target of our automatic assessment, and then elaborate the insights from the interview study with observations from multiple class sessions that are part of a semester-long classroom study. The insights gained from this process then form an empirical foundation for the design of a tool for monitoring group processes. Our goal is to address the real problems faced by instructors in their assessment practices.

Our work is situated within the engineering education community, although concerns related to supporting project based learning are not unique to that community. Project-based learning, especially in courses where students work in groups on real-world problems for and in collaboration with industry sponsors, is commonly believed by educators and administrators alike to have great value for engineering students (Dutson et al. 1997; Adams 2003). These courses are often situated in engineering curricula as capstone design courses that offer students the opportunity to integrate and apply the knowledge they have acquired in more theoretical courses. Moreover, especially because of their connection with industry, these courses offer students the opportunity to move forward in their trajectory from being a student, situated on the periphery of the engineering community, to the core, where they will ultimately function as engineers within that community of practice (Lave and Wenger 1991). Thinking within this socio-cultural perspective on learning, we broaden our consideration of the value students receive from project classes, from a narrow focus on their personal skill set and conceptual knowledge, to a broader focus including their experience of participation, their practice of collaborative knowledge building, and their growth in standing within the professional engineering community.

The value of project courses has been difficult to evaluate from the perspective of disciplinary knowledge gain, which poses challenges for instructors both from the standpoint of formal assessment and, perhaps even more importantly, from the standpoint of being in a position to offer timely, appropriate feedback and support to project teams. The bulk of project learning takes place without the instructor present. Students often flounder out of view of the instructor because they do not know how to begin to construct their own knowledge or effectively play their challenging role as engineers within a team. To support student learning, instructors need to assess student learning processes and outcomes throughout the semester, rather than restricting assessment to final product outcomes. Because the instructor does not direct most of the learning in project classes, the instructor is often unaware of what students have learned, or sometimes failed to learn, both in terms of technical skills and professional skills. Because the final grade is often based on the quality of the product and on the self-reported functioning of the team, students may chose to hide problems from the instructor in an attempt to achieve a better grade, which may ultimately result in poor learning outcomes for many of the students involved.

As illustrative examples of the types of scenarios we are referring to, consider the following two stories collected as part of an interview study we describe in greater detail within this chapter:

One type of frequently reported problem was that students' contributions vary greatly within a team. This is a typical problem in groups referred to as social loafing or free riding (Karau and Williams 1993). This first story is from an instructor, referred to as K, who taught a graduate level capstone design project class. To gain insight into how the groups were functioning, K asked the students to do a peer evaluation in the middle and at the end of the term. To K's surprise, group C rated one member of the group, referred here as Brian, much lower than the rest of the team. K had not anticipated Brian's low score because Brian attended the weekly group meetings in which the group reported their progress to K. Members of group C had covered for Brian during the face to face group meetings with the instructor. K expressed his dismay at not having had enough insight into the group work processes occurring outside the weekly meetings when he observed the group directly to have been able to detect the problem sooner.

Another type of problem reported by a number of instructors was when groups completed their project through a divide-and-conquer approach, i.e., by splitting the project but not communicating with each other while they were working. While this might be an effective way to get work done, it shortchanges the group learning process, removing valuable opportunities for students to learn from one another. In one such situation, S, an instructor for an undergraduate level course on web design, had a pair of students who were working on a project related to building and critiquing websites. The group submitted all the required materials at each of the milestone points. Therefore S did not see any problems with the group until they turned in the final report and product. The report was a conglomeration of two obviously separate parts, each with its own distinct writing style. As in the earlier story, this problem occurred because the majority of group work was done outside of the instructor's view, and the instructor did not have enough insight into the group processes to detect the coordination problems when he could have intervened.

The need is clear for technology to support the important in-process assessment efforts of instructors. To this end, in this chapter we present our work in progress towards the development of a computer-based automatic assessment framework for providing project course instructors with more insight into the group processes within the project teams they are overseeing. We begin with a review of the literature on group processes in project teams, which forms the theoretical framework for our investigation. Next we describe an interview study, which was used to provide an evidence base for the design of our assessment framework. We then describe our process for developing instruments for human application of this assessment framework and subsequent construction of an annotated corpus of project team interactions for use in building prediction models for automatic assessment. Finally, we present some work to date on the process of making the assessment automatic, and conclude with discussion of current directions.

14.2 Theoretical Framework on Group Processes in Project Teams

Group projects are common in instructional contexts because of their many advantages such as tackling problems that are large and complex, assembling broader skill sets from diverse groups (Pea 1993), and offering opportunities to learn from other students (Salomon 1993). Nevertheless, accomplishing a successful group project is not by any means a trivial matter. Process losses (Steiner 1972), which interfere with group performance, may arise for various reasons. For example, conflict may develop between team members (Faidley et al. 2000), some members may engage in free riding behaviors (Karau and Williams 1993), and groups may have difficulty coordinating individual contributions (Strijbos 2004). Because groups do not always function in an ideal manner, it is valuable for graduate and undergraduate programs include project courses to offer students contexts in which they have the opportunity to learn the social skills required for working together in a group under the supervision of an instructor who acts as a group facilitator.

Although the instructors' guidance helps students in overcoming some of the troubles that occur during group work (Hmelo-Silver 2004; Meloth and Deering 1999; McGrath 1984), instructors may have difficulty discerning when support is needed because much of the group work is done when instructors are not present. Sometimes the problems are intentionally hidden behind the well functioning part of the group. Therefore, instructors may miss crucial opportunities to offer support and may not notice group trouble until the problem escalates. Our goal is to enable such episodes to be detected and addressed in a timely manner by providing instructors with insight into group processes as they unfold over time. By being made aware of the blind spots in their understanding of group processes, which is based on limited first-hand exposure to the actual processes occurring in student groups, instructors will be in a better position to diagnose where the problems occur and provide detailed and timely feedback to the groups and individual students alike. In order to address such instructors' needs, we must first identify which processes instructors are most interested in. Although instructors report specific instances of detecting breakdowns in group processes when the groups are having trouble, the list of problems can be more effectively approached when placed within an established conceptual framework. Therefore, we begin with a review of the group process literature to identify existing frameworks.

The literature investigating the connection between group processes and outcomes of group work is both vast and diverse in focus, covering such factors such as communication, coordination, conflict, and conformity (Faidley et al. 2000; Fussell et al. 1998). However, the specific question of which group processes instructors need to be more aware of has not been addressed directly. Usually, only one or a small number of processes are examined to see the effect of those specific processes on group work outcomes. For example, Drach-Zahavy showed that when group members engaged in the processes of exchanging information, learning together, and negotiating over their respective ideas, they displayed increased team innovation

(Drach-Zahavy and Somech 2001). From a different angle, a study by Hackman suggests three types of process criteria closely related to effectiveness of group work: adoption of effective performance strategies, contribution related to unique knowledge and skill, and contribution with respect to level of effort (Hackman 1987).

Some researchers have presented an ontology or a categorization of group processes, rather than focusing on individual processes. However, the categories are typically characterized on a level that is too abstract for our purposes. One example of such a categorization is the separation of processes into task functions and maintenance functions (Fussell et al. 1998; Hackman 1987; Rousseau et al. 2006). Task functions are operation related activities that are performed by group members. Under this heading, researchers have identified behaviors such as managing differences in expertise, skill, and other human resources in order to maximize effectiveness. In contrast, the maintenance functions are overt actions and verbal statements displayed during the team interaction processes that are used to build and maintain cohesion in the group. Under this heading, researchers have examined processes such as openness, trust, and smooth interpersonal relations (Gladstein 1984). In addition to the operations at the group level, such as task and maintenance functions, Gladstein presents a coarser grained categorization of processes, separating them into processes that operate at the organizational level and those that operate at the internal group level.

These studies offer a big picture view of the relationship between certain group processes and outcomes. However, the selection of processes that have been investigated has been guided by a variety of specific interests, different from those that underlie our work. While we can use this prior work to construct an inventory of potentially valuable processes that instructors might be interested in, this literature alone does not identify the most important types of processes from the standpoint of specific instructor concerns. Furthermore, while constructs such as task function versus maintenance function are useful for theory building, it is not clear what the utility would be in organizing processes into task versus maintenance functions as part of our design work if it turns out that instructors do not think in those terms.

14.3 Study 1: Interviews with Instructors

The end goal of our research is to develop technology that is capable of automatically generating periodic summary reports for instructors of project courses, which display representations of important group processes in order to enable instructors to catch problems at an early stage. For example, an automated procedure may be able to identify free riding in a group by monitoring the authorship of reports, postings on discussion boards, or attendance at group meetings. The automated process is desirable in that the group processes could be tracked unobtrusively and even anonymously in order to protect the privacy of students as much as possible. The summary report can be used by instructors to detect and diagnose potential problems before it is too late to intervene, or it could even be used to alert the students that something is not happening as expected. Recent advances in automatic collaborative learning process analysis (Donmez et al. 2005; Joshi and Rosé 2007; Rosé et al. 2008) bring the vision of developing such a tool within practical reach. That technology has proven capable of detecting important conversational events that are indicative of successful group learning in highly controlled settings over short periods of time.

Several research questions must be addressed before such a summary report can be built, especially given that current research on group processes has not addressed the question of diagnosing group problems automatically from the instructor's point of view. Using an interview study methodology, we addressed the following two important questions that provide a foundation for our development effort: (1) What do instructors want to know about their student groups? (2) Is the desired information observable, and can it be reliably tracked by human annotators? In this section we describe an interview study, in which we conducted interviews with instructors and used transcripts of our discussions with them in order to develop ten specific assessment categories.

We began our interview study with two focus questions. First, we asked about the problems that instructors face as they attempt to diagnose problems in group work. Secondly, we probed instructor conceptualizations on how to categorize the observations they make and desire to make into general assessment categories in order to reduce the range of reported issues, difficulties, and reported practices into a manageable list. This section presents the method we employed along with the results from the study. Taking a user centered approach, we interviewed our target users, namely the instructors who teach project courses. Using a grounded theory based analysis approach, we conducted an iterative coding process that resulted in the development of five pairs of assessment categories in addition to a list of indicators that instructors mentioned relying on to make those assessments in their current practice. During the interviews, we found a great deal of overlap in the types of problems instructors mentioned, which demonstrates a certain consistency in the felt gap between the instructors' perceptions of the student groups and reality.

14.3.1 Method: Data Collection and Analysis

Interview data was collected by a team of three interviewers who ran nine focused interviews with instructors. All the instructors had taught at least three university level group project courses. These instructors, all from the same university, included two from design, two from the social sciences, and five from engineering. The interviews lasted from 30 min to an hour. Due to technical difficulties, only six of the nine interviews were recorded; and all of the six recorded interviews were transcribed for further analysis, which included three from engineering, two from the social sciences, and one from design. For the other interviews that were not able to be recorded, detailed notes were taken. We acknowledge then that our analysis will be somewhat biased towards concerns of engineering project course instructors. However, as we mentioned earlier in the introduction, engineering project based learning is our primary concern, and the other data was mainly collected for comparison purposes, in order to get a sense for the generality of the findings. Nevertheless, we stress that we are not making a strong generality claim for project based learning across disciplines.

Background questions about the course included types of projects the students worked on and characteristics of the students who participated in the courses. The purpose of the background questions was to get a sense of the context of the group work the students were doing. Next, instructors were asked to describe the syllabus for their course including the course requirements and how they assigned grades. This information revealed what instructors regarded as important and what they wanted students to learn from the course. Questions about syllabi lead to specific stories of procedures used to assign grades as well as methods for peer evaluations. Then, we asked questions about instances when problems arose in the teams. These stories included details about the cause, detection, and solution of problems. Lastly, interviewers explained to the instructors that these interviews were meant to inform a design effort for a reporting system, or summary tool, that would be meant to help them get more insight into group processes. After explaining the objective of the summary report, we asked a question related to what instructors would want from such a report.

To guard against missing important details, at least two of the three researchers conducting the interviews were present for each interview. After each interview, the researchers typed up their notes and discussed them. After every two or three interviews, all three interviewers got together to consolidate the identified themes based on their respective interview notes. The multiple meetings allowed the interviewers to balance the desire to discuss the content while it was fresh in their minds with the competing desire to base conclusions on deep reflection and comparison across interviews, which often lead to revision of earlier interpretations and the emergence of new themes. Comparing is important since it often leads to revision of initial interpretations, and sometimes to the emergence of new themes. We employed this iterative process in order to obtain meta categories of group processes that instructors look for as they evaluate student groups. From this iterative process, three meta assessment categories of learning, process, and product emerged.

Next, to verify that the meta categories sufficiently covered all the data, the six recorded interviews were transcribed and segmented into sentences. For the six interviews, this yielded a total of 2,320 sentences. The segmented sentences were then coded for further analysis, which we refer to as "assessment category coding". For this assessment coding stage, we selected the sentences related to what instructors wanted to know about the student groups. We excluded sentences on background information, rephrasing of interviewer questions, elaborations meant for clarification, and greetings. Next, to differentiate group process types within the three sets, two coders assigned short descriptive labels consisting of 3 ~ 5 words to the sentences identified as belonging to the three main categories. The short labels were grouped to form 15 detailed categories. The resulting hierarchy of finer grained categories that emerged is displayed in Fig. 14.1, which is shown in Sect. 14.3.2.

Fig. 14.1 Assessment categories

Three rounds of coding occurred during the assessment category coding stage. In the first round, the sentences were each assigned to one of the three meta assessment categories in order to see if they could be reliably differentiated and also to see how much data each of the meta categories covered. In this second round of assessment coding, two coders annotated the 15 detailed categories, which are under the three meta categories to verify that they can be reliably coded. Finally, in the third round of coding, two coders coded for the five pairs of detailed categories under the meta category of "process" because the meta category of "process" is of interest for the purpose of gaining insight into the group processes.

After "assessment coding", we conducted another round of coding, which we refer to as "evidence coding". For each of the five pairs of process assessment categories, we went back to the transcripts and identified the indicators that the instructors mentioned using in order to evaluate them. For example, the extent to which students were each able to articulate what aspects of group work they were taking ownership of was used as evidence that students were equally dividing up their work.

As am informal sanity check, we verified through inspection of the detailed notes that we collected that the three instructors whose data was not transcribed also mentioned the same types of categories, but were not included in the count due to unavailability of transcripts.

14.3.2 Results

Among the three coarse grained assessment categories of learning goals, process, and product, process is of greatest interest for two reasons. First, group difficulties can be revealed through interaction processes that display such things as amount of effort offered by group members or characteristics of group dynamics. On the other hand, learning goals set by instructors or the resulting group products do not show where and when the students are having difficulty while doing group work. Although in real work settings, it is mainly the success of the final product that matters, instructors regard the process to be important for the purpose of giving students the opportunity to learn. By influencing the process, instructors have the opportunity to enhance both the learning experience as well as to facilitate the accomplishment of a higher quality product. By the time the product has been produced and the learning objectives of the course have been accomplished, it is too late for the instructors to intervene. Finally, processes are also a more appropriate focus for a tool that is meant to be general across multiple disciplines. The same group processes are relevant in teamwork within the domains of design, behavioral science or engineering. However, the learning objectives as well as the group products differ across disciplines and within the same discipline.

The importance of looking at process was evident in the data as well. Instructors mentioned assessment categories under process more often (70% of the instances) than under learning goals (15%) or product (15%). In addition, the number of more detailed assessment categories under the general heading of process (10) was higher than those mentioned under learning goals (3) or product (2) as seen in Fig. 14.1. Given the importance and interest, we focus on the meta-category of process rather than learning goals or product in the rest of the chapter. Note that the ten assessment categories under the general heading of process can be paired into corresponding individual and group level assessments. Individual assessments relate to an individual student in isolation, whereas group level assessments relate to students in connection with their group or in comparison with their team members. Although the meta category of learning goals and products can be also divided into the individual and group level, we did not divide those categories further because our focus is on

Table 14.1 Five pairs of processes assessment categories

Individual	Definition (# of instructors mentioned)	Group	Definition (# of instructors mentioned)
Personal goal setting	Making individual plans for the next steps (4/6)	Group goal setting	Making team plans for the next steps (4/6)
Personal progress	Fulfilling personally stated goals through producing work (5/6)	Group progress	Fulfilling group goals through producing work as a group (4/6)
Knowledge contribution	Taking initiative to use knowledge or skill (1/6)	Group knowledge building	Exchanging skill, idea, or conversation which leads to learning & project advancement (3/6)
Participation	Being involved in work (6/6)	Division of labor	Contributing work for the group relative to other group members (5/6)
Team player	Attitude toward interacting with the team (5/6)	Interpersonal dynamics	Interaction within the team due to personality & relationship (5/6)

processes rather than learning or group products. Table 14.1 shows the definitions of the five pairs of process assessment categories.

For the three rounds of assessment category coding, we achieved the following kappa values. In the first round of coding where two coders coded the three meta categories, we calculated a kappa agreement of 0.88 between two coders over 20% of the data, an acceptable rate of agreement. For the second round of coding where two coders coded the 15 categories, a kappa value of 0.72 was achieved for the 20% of data. For the third round of coding where two coders looked at the five pairs of assessment categories under process, the coders coded 20% of the data on the five assessment categories and achieved a kappa of 0.90. After calculation of the kappa in each round, disagreements were settled by discussion among the coders.

As mentioned above, after categorizing the list of processes that are used for evaluating group work, we conducted "evidence coding" to see what pieces of evidence are currently used by instructors in order to track the five pairs of categories under the meta category of "process". The list of evidence mentioned by the instructors contained both directly observable and inferable evidence. Directly observable evidence is most visible, and therefore more straightforward to track both from a human judgment standpoint and from a technical standpoint when we move into automatic assessment later in the article. For instance, the number of postings on a message board is directly observable, but inferences from conversations that take place at group meetings, even when these details are shared with instructors, are not. Although, inferable evidence is harder to track, it is as frequently mentioned as the directly observable evidence. Therefore, methods of detecting inferable evidence should be investigated in future studies. We now present each of the five pairs of process assessment categories that we looked at in more detail along with the pieces of evidence that the instructors mentioned in connection with each of the categories.

Related to the idea of looking for evidence on which to base judgments associated with the assessment categories discussed above are the observations instructors mentioned making of the eventual fate of groups where there were problems related to those assessment categories. These types of observations, which are detailed in Table 14.2, serve as negative evidence when issues have been left undetected and therefore unaddressed for too long. Ultimately, a major long term objective of our work is to assist instructors in catching problems at an early stage so that these types of observations become much rarer than they are.

Now we discuss each one of our assessment categories in detail. We offer a detailed description of each group level and individual level subcategory and discuss the evidence that instructors mentioned using to make their assessments.

14.3.2.1 Personal Goal Setting and Group Goal Setting

The first pair of assessment categories is personal goal setting and team goal setting. Goal setting involves making concrete plans for the project's next steps. For instance, instructors assessing personal goal setting might look for students

Table 14.2 Study 1 results: Instructor identified needs and problems

Assessment categories	Example processes wanted by instructors	Example problems observed
Personal & group goal setting	Selecting own research methods and putting together own research plans	Spend too much time having meetings without productive results
Personal & group progress	Steering and controlling process, check the accomplishments. Keeping track of where they are in the project	Not meeting production goals, bottlenecks occur when part of the team is not delivering
Knowledge contribution/ building	Members sitting all together, physically close and being in constant communication with a tight feedback loop	Produce reports that are not united and have clearly separate sections
Participation/ division of labor	Each member contributing and presenting their work to terms	Some people carried by their team members and the one who worked complain
Team player/ interpersonal dynamics	Not having trouble working together and collaborating with each other	End up with a dysfunctional team where the members are not even talking to each other

selecting methods and putting together a project plan with explicit milestones. For group goal setting, instructors might examine whether the whole team is setting an appropriate goal. Having all of the team members buy into the same vision is important for group goal setting as well.

To assess student's personal goal setting, some instructors observed whether a student produced a list of tasks to accomplish using a schedule or activity charts. Others looked at publicly stated goals that each student made during weekly meetings. Some instructors were more explicit and required students to submit lists of tasks as well as the time spent on each task. The frequency of the submission of such lists varied from weekly to monthly depending on instructors' preferences. In addition to personal goals, instructors looked for team goals. One instructor mentioned that to see whether a team had a goal, she observed group meetings. If the meeting lasted too long or did not have any explicit agenda, that indicated the team should have made more specific plans. Instructors also looked at schedules produced by the group that show dependencies between their tasks to see whether groups are doing an effective job of coordinating across activities.

14.3.2.2 Personal Progress and Group Progress

In addition to suggesting and providing goal setting help for students, instructors followed up to see if the students were fulfilling their stated goals. Instructors mentioned that they observed whether students fulfilled promises they made, whether students steered and controlled the process of their work, and whether they checked their accomplishments along the way. For group progress, instructors checked whether groups explicitly tracked progress towards milestones agreed upon as a group. To assess personal progress, instructors looked at schedules to see whether planned items were finished on time and which action items were accomplished. For group progress, instructors observed scheduled team meetings, which varied in frequency from once a week to three times a semester depending on the course. In these meetings, instructors looked at students' presentations on what they had done so far as a group, and at the team's progress by examining the list or resolutions made. In addition to the meetings, instructors looked at midterm and final presentations for similar assessments.

14.3.2.3 Knowledge Contribution and Group Knowledge Building

Knowledge contribution and group knowledge building is the next pair of assessment categories. At the individual student level, instructors observed students taking initiative to use their unique knowledge and skills in doing group work. In addition, at the group level, instructors looked for evidence of group members mentoring other students on skills, sharing ideas, and engaging in meaningful conversations that may lead to learning and project advancement. Note that not all conversational contributions would be considered instances of a knowledge contribution in the truest sense. Evidence of reasoning, either of the student's own or other's ideas/knowledge, is necessary.

Overall, the pieces of evidence instructors used to assess knowledge contribution and group knowledge building were not very concrete or direct. In general, instructors looked more for evidence of breakdowns of these processes, rather than positive evidence of the occurrence of these processes. For example, instructors mentioned that one sign of trouble is when students come to talk to them about the absence of communication with other students in their group. Note that the same piece of evidence can be used in connection with other categories as well. For instance, absence of communication can be used as an indicator for the participation category. However, evidence such as attendance at group meetings can only be used for participation and not knowledge sharing. Knowledge sharing involves exchange of ideas and skills, and thus likely requires traces of communication as pieces of evidence. In addition to communication, another interesting indicator that instructors used for knowledge contribution was when the overall productivity of a group was low. In addition, an unintegrated work product, such as the patched report described in the second story which was illustrated in the introduction of this chapter, indicated trouble. To assess knowledge contribution and group knowledge building, it might be possible to infer knowledge building breakdowns from the available traces of group work. For instance, a low number of discussion threads initiated by students or a low number of replies by students on a group message board may indicate that communication is not active. Other sources of information such as exchange of emails or number of group meetings could be used as additional indicators.

14.3.2.4 Participation and Division of Labor

Another pair of assessment categories is participation and division of labor. To assess participation, instructors observed whether each student contributed to the group effort or whether some students were not working. If a student was not working in an obvious way, such as not attending group meetings or classes, instructors could easily detect such instances. However, instructors were also concerned with potentially not knowing about students who seemed diligent, but in reality were slackers. In addition to participation, instructors wanted to know each student's contribution to see whether the work is done by only a subset of the group members.

The importance of participation was articulated by all six of the instructors. To observe the assessment category of participation, instructors examined a variety of sources including self reported work logs, peer evaluation forms, and group message boards. From these various sources, they looked at student attendance in classes or group meetings, number of hours worked, number of action items accomplished, number of messages posted in group message board. Indicators used for the assessment category of division of labor also came from the same sources as the individual assessment category of participation. Instructors used the same indicators such as number of hours worked and compared to those of other group members to see whether the distribution of work is equivalent among the members.

14.3.2.5 Team Player and Interpersonal Dynamics

The last pair of assessment categories is engagement and interpersonal dynamics. Engagement is attitude towards participating in doing group work. Instructors looked for students that were dedicated, emotionally invested and loved their project. One instructor noted that when she listened to students presenting, she looked for engagement as the following quote demonstrates: "I think with experience I see when somebody is really honestly engaged with the work or whether they are faking it. So that's the first thing I look for." The group level assessment category relating to attitude is interpersonal dynamics, which is interaction within the group resulting from personality and relationships. Instructors observed the group chemistry such as whether students were having difficulty getting along.

To assess students' level of engagement, instructors observed their behavior. For instance if a student created posts and avidly replied to other students' posts, instructors inferred that the student was engaged. Another instructor noted that when students were enthusiastic about their project, they came to the instructor or other team members willingly for more work. Also, when the instructors saw students out in the field building things rather than just browsing the web looking for more information and putting off the "Getting your hands dirty" part of the project, they inferred those students were actively involved in the project. Instructors stated that although group dynamics is an assessment category they would like to gain more insight into, currently no good indicators exist other than spending time and being involved with the team. However, researchers have inferred team dynamics by looking at the type of language used in the team. For instance examining the usage of positive words versus negative words used by students in their correspondence (Pennebaker et al. 2008) may signal a particular pattern of group dynamics. Another is to observe the cohesiveness of conversation by looking at the words used in message board or documentation produced by the group, where a low degree of cohesiveness may signal conflict between team members (Dong et al. 2004).

14.4 Study 2: Project Group Observations

Given the five pairs of assessment categories that instructors look for when evaluating group projects, the next step involves tracking those categories. Before tracking the assessment categories using machine learning technology, we first verified that the information desired by the instructors is observable and can be reliably tracked by humans. The second study addresses this issue by having two researchers observe project group meetings and manually evaluate the targeted assessment categories. The end product of this process was both an instrument for human assessment using the ten assessment categories that came from our interview study as well as a corpus of hand annotated meeting recordings to use in the technical development effort.

14.4.1 Method

The course that provided the context for our data collection effort is a graduate level engineering course that was offered in spring 2008, where the students work on one big project sponsored by a client. Four subgroups were formed in order to carry out the project. The fact that the subgroups were part of one larger project would need to be considered if one were to analyze each of the subgroups' tasks. However, because our research focused on the group work process rather than the task themselves, we do not expect the relationship between the groups to affect the results of the analysis substantially. Being a project oriented class, a major component of the grade assigned by the instructor is based entirely on their productivity, and this portion of the grade is explicitly indicated by the instructors, separate from the part of the grade related to the quality of the result. There were two instructors and 22 students in the class. Various types of data were collected in this class, including messages on discussion boards, reports, and weekly work logs from each student. However, this information is not enough to address the gap between what instructors would like to know about the groups they are overseeing, and what they actually see. In order to get a more specific picture of what information instructors are missing, we instrumented the course in order to collect extensive observational data from the groups. Specifically, we collected audio recordings of group meetings as well as video tapes of classroom activities. In order to develop a reliable instrument for humans to use in making assessments related to the ten categories, we had two researchers sit in during weekly group meetings for the first 5 weeks of class. Overall, the development effort was based on approximately 10 h of meeting data.

More specifically, two researchers sat in during weekly group meetings and evaluated the five pairs of assessment categories identified in the first stage of our work. Group meetings were chosen as the target of our observation since the bulk of group project work is accomplished during the group meetings, although instructors are not able to attend group meetings due to time constraints. The two sets of assessment categories scored by the two researchers were used to calculate a reliability measure, which would show whether the assessments can be made reliably from these observations. In observing the group meetings, the researchers remained uninvolved in the group meetings, as "flies on the wall". The main goal of this method is to observe the environment and the social interactions as they occur without influencing the participants. In order to achieve this goal, the students were assured that their grade would not be affected in any way due to the presence of the researchers and that there would not be any communication regarding the group meetings between the researchers and the instructors.

The scoring of the ten categories was conducted in the following way. For each of the ten categories, we constructed statements that described the positive student behaviors associated with each given process, which were drawn from our evidence coding discussed above. For example, for the category of "interpersonal dynamics", one of the statements is, "Is everyone's opinion taken seriously without being ignored? Is there an attitude towards valuing everyone's suggestions?". For each statement, observers answered "yes" or "no". Then, based on the observations made in response to these statements, scorers assigned an overall score for the "interpersonal dynamics" category with a number between −2 and 2. For instance, answering "no" to all the statements described in the category would result in an overall score of −2, whereas "yes" to all the statements would result in a score of "2". The range has both negative and positive numbers so that the scorers can easily map negative behaviors to negative scores and positive behaviors to positive scores. Five numbers can also capture the difference between behaviors sufficiently as used in many grading systems (e.g., A ~ F). The observations were conducted weekly, and scores were assigned for each of the four groups for each group level category, and for each of the 22 students for each student level category.

14.4.2 Results

The reliability of the coding scheme for the ten categories was verified by calculating the correlation between the scores assigned by the two researchers, which was 0.81 for the group level categories and 0.64 for individual level categories. Several observations were made during the group coding process. First, assessing all ten categories was too much of a burden for the coders. One possible direction for the future that we are considering is to limit our focus to individual level assessments. We expect that even if it is only the individual level categories we are able to offer more insight to instructors about, by seeing which individual are in trouble, teachers may be in a better position to allocate their time effectively, and to do more detective work to determine whether the individual performance would affect the group in each problem case.

A second observation we made was that although the coders felt that they had a harder time assigning the goal setting and group progress assessment categories than knowledge building, division of labor, and interpersonal dynamics, the correlation of scores between the coders were actually higher for goal setting and progress ($r=0.92$) as opposed to the latter three ($r = 0.75$). This indicates that the confidence that instructors may feel about their judgments related to the different assessment categories may be unreliable.

A third, and possibly most important, observation was that depending on the type of the meeting, the type of assessment categories that were observable differed. The two main types of meetings that the students held were administrative meetings where most of the discussions were related to handling administrative matters such as scheduling and work assignments, and work meetings when the bulk of the meeting time was spent doing actual work such as building conceptual framework for their project. For the administrative meetings, goal setting and progress assessment categories were easier to observe, whereas almost none of the knowledge sharing occurred, making it hard to assess knowledge building assessment categories. However, for the work meetings, knowledge building was easier to assess where as the goal setting and progress behaviors were rarely discussed. For purposes of automatic assessment, we will have to explicitly take the type of meeting into account when using data collected from the meetings to make assessments, taking into consideration which assessment categories we can get a reliable assessment for depending upon the type of meeting.

We refined the assessment instrument during the first third of the course. In the second third of the course, we used the instrument to record assessments for scheduled group meetings. In the final third of the course, the observers continued to attend scheduled group meetings, however the frequency of such meetings dropped considerably during the final third of the course until they eventually ceased altogether in favor of impromptu small group meetings that occurred in on as-needed basis.

14.5 Study 3: Developing Technology for Automatic Monitoring of Group Processes

Given that the process assessment categories are observable and traceable by human annotators, our next research challenge is identifying methods of automatically tracking those processes and displaying them to instructors so that instructors can gain more insight into the group processes. This is not the first such effort. For example, Joshi and Rosé (2007) found that machine learning techniques applied to chat logs from collaborative learning discussions were more accurate at ranking how well student groups learned together than humans observing the complete chat transcripts. Similarly, McLaren et al. (2007) have taken a similar approach in the Argunaut project where they have investigated types of assessments instructors wanted to make about online learning groups and used machine learning to do that analysis. Based on prior work such as this, we expect that we can use machine learning technology to track important group processes by leveraging the data we are collecting during the class. For instance, a report where individual students' weekly statistics are displayed could be given to the instructor if this could be computed from collected data. To this end, in our previously published work (Rosé et al. 2007), we have explored how we can make automatic assessments from message board data. This exploration was based on data collected in a previous semester. In this section, we present a brief overview of findings from those initial explorations with message board data prior to the development of the assessment framework described in this chapter. We then report our work to date automating application of the assessment categories from our framework through automatic speech processing.

14.5.1 Automatic Assessment from Message Board Data

We chose productivity as an outcome measure for our initial feasibility test because we had access to an instructor assigned productivity grade assigned at three separate times in the semester for each of the students in the 2006 course. Using the discussion board data, we collected a total of 1,157 posts. We divided the data into collections of posts posted by an individual student on each specific week of the course. Altogether we assembled 476 collections of such posts. From this raw data, we extracted a number of linguistic features using TagHelper tools (Donmez et al. 2005; Rosé et al. 2008), which can be downloaded from http://www.cs.cmu.edu/~cprose/TagHelper.html. These features included stemmed unigrams (i.e., single word stems), and bigrams (i.e., pairs of word stems occurring adjacent in the text). We did not include words that occurred less than five times in the corpus or typical function words such as prepositions or determiners. In addition to this we included features extracted by using Pennebaker's Linguistic Inquiry and Word Count (LIWC) (Pennebaker et al. 2008) as well as other surface features of the conversational behavior such as number and length of posts. Next we used the support vector machine (svm) regression model included in the Weka toolkit (Witten and Frank 2005) to build a predictive model. The SVM regression algorithm is a type of regression learner that learns from data how to weight various features that we provide. We begin by extracting multiple features from the collections of posts that we think will be relevant. We label instances, which are vectors of such feature values, with the instructor assigned grades associated with the week the data is from. The regression learning algorithm then learns weights for the features depending on their predictive value with respect to the instructor assigned scores. Predictions can then be made for vectors of feature values by evaluating the learned linear function.

Because the instructors assigned grades three times throughout the semester, the grade associated with an individual week was the grade assigned to the student for the segment of the course that week was part of. Since student productivity may vary from week to week throughout the semester, we consider this target assessment to be somewhat noisy. However, it is the best objective measure we have for individual student productivity in the course.

Using a methodology in which we train on part of the data and test on the remainder of the data, we have been able to build a model that can make reasonably accurate assessments of productivity (R =.63). Interestingly, some of the most predictive linguistic features were social in nature. For example, we observed that words such as "thanks," "hi" and "please" ranked among the top attributes in the feature space. Because the words are social in nature, one possible explanation is that the groups that are more likely to socialize are more productive. However, this hypothesis must be verified in future studies. While we are continuing to work to improve this prediction accuracy, we believe these preliminary results show promise than an automatic analysis of on-line communication behavior of groups can provide instructors with valuable early warning signs that some groups or certain students within groups require some additional instructor support.

14.5.2 Automatic Assessment from Speech

Using the recordings collected from each student during group meetings, we computed an estimate of activity level for each student using machine learning technology. Average activity level is an approximate measure of the amount of talk that the student contributed during group meetings. Before this could be computed from the speech, the recordings needed to be segmented, and each segment needed to be coded for the amount of speech by the associated student that the recording was of. We chose to segment the speech into 10 s intervals so that it would be reasonable to assume that for most segments, there would be at most a single dominant speaker. That allowed us to utilize a relatively simplistic approach to coding activity level for individual segments. We adopted the following 4-point scale for activity level: 0 - no speech from primary speaker; 1 - primary speaker only does back-channeling, where back-channeling is a way of showing a speaker that you follow and understand their contributions, often through interjections such as, "I see", "yes", "OK", "uh-huh"; 2 - primary speaker speaks holds the floor for less than half of the 10 s; 3 - primary speaker speaks holds the floor for more than half of the 10 s.

We first verified that human annotators could make this judgment reliably from the audio recordings of individual segments. Using this coding scheme, the inter rater reliability evaluated for two coders over 144 segments was 0.78 Kappa. With the reliable coding scheme, a single coder then coded 1,132 segments (distributed evenly across students, from meetings during Phase 1 of the course). The largest proportion of segments was coded as 0, which amounted to 47.5% of the segments. 8.5% were coded as 1, 30.5% as 2, and 13.5% as 3.

Before it is possible to apply machine learning to speech in order to automatically replicate the coding just discussed, each segment of speech must first be transformed into a set of feature-value pairs. The activity level that we are trying to predict from speech is related to "how" the words are spoken rather than the content of the words. Such stylistic aspects of speech are captured by speech prosody. Similarly, other speech applications such as emotion detection are also more concerned with speech prosody rather than the content (Beskow and Sjlander 2000), and thus use features similar to ours. In contrast, speech applications such as dictation software use content related features such as spectral features processed through a speech recognition system. Therefore, the features extracted from speech for our experiment are variations of prosodic features such as pitch, power and amount of silence. A total of 39 prosodic features were extracted for each of the 10 s segments using wavesurfer (Beskow and Sjlander 2000).

With the coded speech data after it had been transformed into a vector representation consisting of the 39 features just discussed, we then evaluated whether it was possible to use machine learning to automatically assign segments of speech to one of the four coding categories mentioned above with high enough accuracy. In other words, the question was whether it would be possible to find patterns within that representation of the speech that would allow for automatic replication of the human coding. Because the size of the dataset was small, we used Weka's SMO (Weka's implementation of support vector machine) learning algorithm (Witten and Frank 2005) because it is known to be able to avoid overfitting better than other machine learning approaches. In order to safeguard against the evaluation results being inflated due to overlap in speakers between train and test sets, we adopted a cross-validation methodology where a model was first trained on all but one student, and then performance was evaluated over the segments of the remaining students. We did this for each student and then averaged across students to compute the performance of 74.26% accuracy.

Knowing the accuracy is not sufficient, however. There is still the question of how accurate is accurate enough. So we then validated our machine coding by evaluating how well it allowed us to achieve a comparable average activity rating for each student relative to the average activity rating we could get from the human four point coding. In order to do this validation, we first computed an average activity rating for each student in each meeting by averaging over the human assigned codes. We then did the same with the computer assigned codes. When we correlated the average activity levels for each student based on human codes with those based on the automatic codes, we achieved a correlation coefficient of 0.97, indicating that we can achieve a reliable estimate of activity level using a machine learning model.

Based on this validation, we concluded that the model was accurate enough to code the remainder of our speech data. Thus, we then trained a model using all of the coded data, which we used subsequently to code the data used in the correlation

Table 14.3 Correlation between activity from speech and observer ratings for each student along dimensions of goal setting, progress, knowledge sharing, division of labor, and the average across dimensions. Note that the group dynamics dimension was left out of this correlation analysis due to lack of variability in scores

	Observer goal setting	Observer progress	Observer knowledge sharing	Observer labor division	Observer average
Activity from speech	.514	−.031	.309	.351	.447

analysis presented later in the paper. For the test data, a recording for a meeting that was submitted in phase 2 was randomly selected for each student. Four students never turned in any recordings, so 18 students' recordings were segmented into 10 s segments. The length of each recording differed due to differences in meeting lengths. The number of segments ranged between 7 min 30 s and 2 h 19 min 50 s in length (45–839 10 s segments), with an average of 47 min in length (282 segments). Using the speech model built in step 3 with the Weka toolkit (Witten and Frank 2005), student recordings from phase 2 were assigned average activity level ratings.

Average activity level ratings are only useful to the extent that they can be used to predict the assessment categories we started with. In Table 14.3 we report correlations between the automatically computer activity level and the assessment categories as they were assigned by observers in the group meetings. Unfortunately, we were not able to compute correlations for the Group Dynamics dimension due to insufficient variability in the instructor ratings on that dimension. Note that our analysis focuses on individual level assessments since the indices we extract from the speech are from individual recordings. Because we are comparing observer ratings, we inserted the word "observer" on the column heading for clarification.

While clearly there is much room for improvement, we have only begun to scratch the surface in terms of what can be detected in the speech recordings collected from group meetings. In our current work, we are annotating the speech for explicitly displayed reasoning and instruction and plan to use automated indicators of these processes to further elaborate our automatic assessment model.

14.6 Discussion

In the three studies presented in the paper, we have presented the assessments that the instructors need to make in order to diagnose group problems as well as the feasibility of performing the identified assessments both by human annotators and using machine learning technology. One issue not addressed thus far is how best to communicate this information to instructors. This problem is somewhat related to issues addressed in the conversation visualization community (Shneiderman 1992; Smith and Fiore 2001), because our work aims to display information inferred largely from conversation data. However, because our work does not focus on displaying a representation of the conversations themselves, but rather what is learned about the groups using that data, our work is closer in spirit to other work such as the Wattle tree (Kay et al. 2006) and Group Awareness Widgets (Kreijns et al. 2002).

In this section, we present our vision for the summary report that the instructors can use to categorize and diagnose complex problems such as those we identified in our interview study. Although the system has not been implemented yet, we included two sketches (Figs. 14.2 and 14.3) to illustrate the vision. Specifically, we describe how K, mentioned in the introduction of the chapter, could have used a sample summary report to diagnose problems. Imagine that at week 4, K looks at the summary report, which he uses approximately once a week to see if any students need guidance. A sample summary report shown in Fig. 14.2 shows group C's progress on the five group level assessment categories.

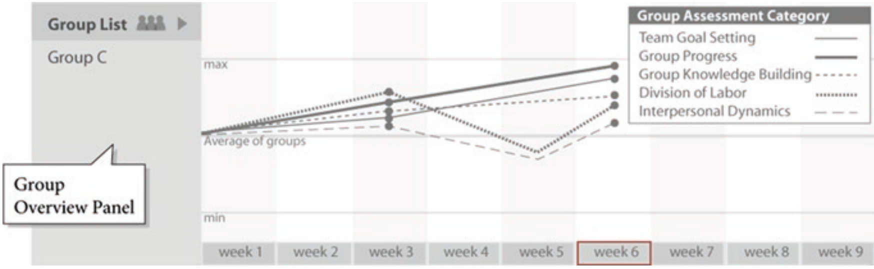

Fig. 14.2 Graphs displaying the five group level assessment categories over time

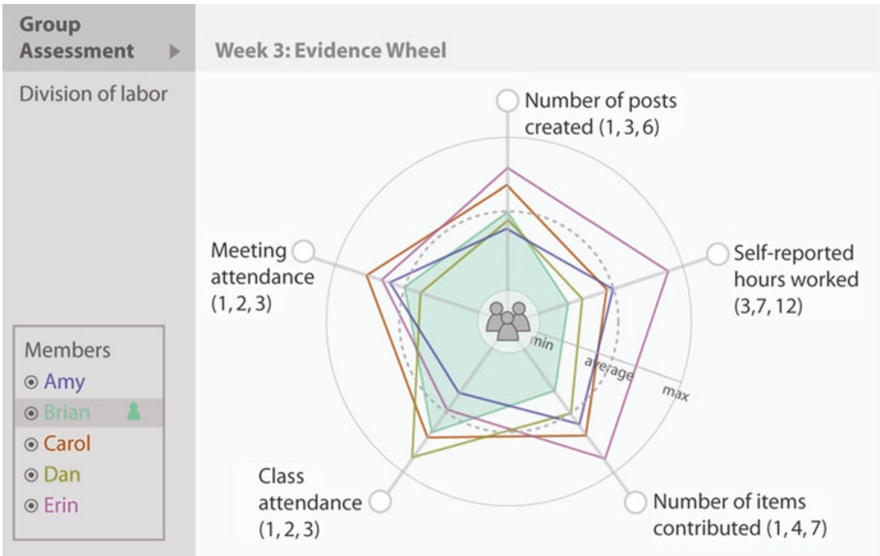

Fig. 14.3 Evidence wheel for showing Brian's below average contribution in the shaded octagon

K can select any of the group assessment categories to look closer at the indicators used to measure the categories. K decides to examine the division of labor in more detail because he observes that it is one of the declining categories. In order to look at the data in more detail, K can use the evidence wheel, which shows the quantitative values of evidence used to compute each assessment category. As seen in Fig. 14.3, an evidence wheel for the assessment category of division of labor could be displayed for group C. Notice how the evidence wheel contains three concentric circles, which show the minimum (innermost), average, and maximum (outermost) values computed from the values of all group members. Brian's octagon is the inner most shape shaded in gray. The evidence wheel shows how an indicator, such as "number of posts created," compares across selected students by the relative position of students on the radius of the circle. Each indicator also lists the minimum, average, and maximum values below its label. The points for an individual student are connected to form a pentagon as shown in Fig. 14.3. Brian's pentagon is shaded, and by looking at each of the vertices of the shaded pentagon, K can see the relative quantitative values of each indicator for Brian. Because Brian's self reported hours worked and number of items contributed are close to the minimum, K sees this as indicative of a problem. At this point, the instructor can take further action, such as having a meeting with the group to discuss the concern and work out a plan of action.

14.7 Conclusions and Future Work

The goal of our project is making group work processes that are normally hidden or implicit more obvious and explicit. By formalizing, quantifying, and displaying indicators in the summary tool, we hope to support instructors in carrying out the assessment processes that are already part of their general practice and to improve the quality of these judgments by bringing more data to bear on their interpretation process so that their assessments will be more reliable. We presented a mixed methods approach where we combined an interview study to gain insights from instructors and a classroom study to form an empirical foundation for a design of our tool. Researchers who are studying group processes could also adopt our approach to take advantage of expert insights that may otherwise be missed. Establishing such common ground with practitioners in the field is important in order to build a system that would be useful to them.

The work presented in this paper forms a foundation for tracking important group processes, however, further work needs to be done. The immediate next steps, as illustrated in the previous sections, include comparing the instructor assessment with those of the researchers to identify less visible processes to the instructors and building a prototype tool using the collected classroom data to predict the assessment categories using machine learning technologies. Although increasing the accuracy of judgments can be accomplished to a certain level, given the current technology and the nature of the problem it would be difficult to achieve a 100% accuracy. Therefore, another issue that should be considered is displaying the level of certainty of the predictions so that instructors can make appropriate decisions.

One limitation of our current work is that our interview study only involved nine instructors. Furthermore, the number of instructors we had from each field was too small to do a systematic comparison across fields to determine whether there are systematic differences between group project issues across fields. Secondly, all of our data is from the instructor's point of view and not from that of the students. It would be interesting to follow up with interviews of students to identify what they see as lacking in the support their receive from their project course instructors.

Acknowledgments We are deeply grateful to the instructors that participated in our interview study as well as the students and instructors who participated in the classroom study. Without their patience and participation, this work would not have been possible. This research was supported in part by NSF EEC grant number 064848.

References

Adams, R. S., Turns, J., & Atman, C. J. (2003). Educating effective engineering designers: The role of reflective practice. *Design Studies, 24*(3), 275–294.

Beskow, J., & Sjlander, K. (2000). Wavesurfer–an open source speech tool. In *Proceedings of ICSLP*, Beijing, 464–467.

Chen, M. (2003). Visualizing the pulse of a classroom. In *Proceedings of international multimedia conference* (pp. 555–561). Berkeley: ACM Press.

Dong, A., Hill, W. A., & Agogino, A. M. (2004). A document analysis technique for characterizing design team performance. *Journal of Mechanical Design, 126*(3), 378–385.

Donmez, P., Rose, C. P., Stegmann, K., Weinberger, A., & Fischer, F. (2005). Supporting CSCL with automatic corpus analysis technology. In T. Koschmann, D. Suthers, & T.-W. Chan (Eds.), *Proceedings of computer supported collaborative learning* (pp. 125–134). Taipei: Laurence Earlbaum Associates.

Drach-Zahavy, A., & Somech, A. (2001). Understanding team innovation: The role of team processes and structures. *Group Dynamics, 5*(2), 111–123.

Dutson, A. J., Todd, R. H., Magleby, S. P., & Sorensen, C. D. (1997). A review of literature on teaching design through project-oriented capstone courses. *Journal of Engineering Education, 76*(1), 17–28.

Faidley, J., et al. (2000). How are we doing? Methods of assessing group processing in a problem-based learning context. In D. H. Evensen & C. E. Hmelo (Eds.), *Problem-based learning: A research perspective on learning interactions* (pp. 109–135). NJ: Erlbaum.

Fussell, S. R., et al. (1998). Coordination, overload and team performance: Effects of team communication strategies. In *Proceeding of conference on computer-supported cooperative work, Seattle* (pp. 275–284). New York: ACM Press. 14–18 November 1998.

Gladstein, D. L. (1984). Groups in context: A model of task group effectiveness. *Administrative Science Quarterly, 29*, 499–517.

Hackman, R. (1987). The design of work teams. In J. Lorsch (Ed.), *Handbook of organizational behavior*. Englewood Cliffs: Prentice-Hall.

Hmelo-Silver, C. E. (2004). Problem based learning: What and how do students learn? *Educational Psychology Review, 16*(3), 235–266.

Joshi, M., & Rosé, C. P. (2007). Using transactivity in conversation summarization in educational dialog. In *Proceedings of the SLaTE workshop on speech and language technology in education.*

Karau, S., & Williams, K. (1993). Social loafing: A meta-analytic review and theoretical integration. *Journal of Personality and Social Psychology, 65*, 681–706.

Kay, J., Maisonneuve, N., Yacef, K., & Reimann, P. (2006). *Wattle tree: What'll it tell us?,* University of Sydney, Technical Report, p 582.

Kreijns, et al. (2002). The sociability of computer-supported collaborative learning environments. *Journal Of Education Technology & Society, 5*(1), 8–22.

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. New York: Cambridge University Press. ISBN 0521423740.

Madan, A., Caneel, R., & Pentland, A. (2004). Group media: distributed multimodal interfaces. In *Proceedings of sixth international conference on multimodal interfaces ICMI'04.*

McGrath, J. (1984). *Groups: Interaction and performance*. Englewood Cliffs: Prentice-Hall.

McLaren, B., Scheuer, O., De Laat, M., Hever, R., de Groot, R., & Rosé, C. P. (2007). Using machine learning techniques to analyze and support mediation of student e-discussions. In *Proceedings of artificial intelligence in education.*

Meier, A., Spada, H., & Rummel, N. (2007). Evaluating collaboration: A rating scheme for assessing the quality of collaborative process. *International Journal of Computer-Supported Collaborative Learning, 2*, 63–86.

Meloth, M. S., & Deering, P. D. (1999). The role of the teacher in promoting cognitive processing during collaborative learning. In O'Donnell & King (Eds.), *Cognitive perspectives on peer learning*. Hillsdale: Lawrence Erlbaum Associates.

Pea, R. D. (1993). Practices of distributed intelligence and designs for education. In G. Salomon (Ed.), *Distributed cognitions* (pp. 47–87). New York: Cambridge University Press.

Pennebaker, J. W., et al. (2008). *Linguistic inquiry and word count*. http://www.liwc.net/. Retrieved 24 April 2008.

Rosé, C., Gweon, G., Arguello, J., Finger, S., Smailagic, A., & Siewiorek, D. (2007). Towards an Interactive Assessment Framework for Engineering Design Project Based Learning. *In Proc. DETC* 2007.

Rosé, C. P., Wang, Y. C., Cui, Y., Arguello, J., Stegmann, K., Weinberger, A., Fischer, F. (2008) Analyzing collaborative learning processes automatically: exploiting the advances of computational linguistics in computer-supported collaborative learning. *International Journal of Computer Supported Collaborative Learning.*

Rousseau, V., Aube, C., & Savoie, A. (2006). Teamwork behaviors: A review and an integration of frameworks. *Small Group Research, 37*(5), 540–570.

Rienks, R. J., Zhang, D., Gatica-Perez, D., & Post, W. (2006). Detection and application of influence rankings in small group meetings. In *Proceedings of eighth international conference on multimodal interfaces ICMI'06.*

Salomon, G. (1993). No distribution without individual cognition: A dynamic interactional view. In G. Salomon (Ed.), *Distributed cognitions* (pp. 111–139). Cambridge: Cambridge University Press.

Shneiderman, B. (1992). Tree visualization with tree-maps: 2-d space-filling approach. *ACM Transactions on Graphics, 11*(1), 92–99.

Steiner, I. D. (1972). *Group process and productivity*. New York: Academic.

Strijbos, J.W. (2004). *The effect of roles on computer-supported collaborative learning*. Ph. D dissertation, Open University of the Netherlands, The Netherlands

Smith, M., & Fiore, A. (2001). Visualization components for persistent conversations. In *Proceeding of conference on human factors on computing systems, Seattle* (pp. 136–143). New York: ACM Press. March 31 – April 5, 2001.

Weinberger, A., & Fischer, F. (2005). A framework to analyze argumentative knowledge construction in computer-supported collaborative learning. *Computers & Education, 46*(1), 71–95.

Witten, I. H., & Frank, E. (2005). *Data mining: Practical machine learning tools and techniques* (2nd ed.). Elsevier: San Francisco.

Chapter 15
Analyzing Productive Interactions in CSCL: Collaborations, Computers and Contradictions

Eileen Scanlon

Abstract This chapter reviews a series of case studies taken from research projects conducted in the computers and learning research group at the Open University examining ways to investigate computer supported collaborative learning interactions. The aim of this series of experiments was as part of a research programme directed at developing a better understanding of the way in which technology enables collaborative learning. A range of projects where technology has been used to support collaboration in a variety of settings is reviewed here. These include settings where adults were collaborating on problem solving tasks at a distance (using technologies to support collaboration such as Shared Ark and Kansas), and young people using mobile technologies and collaborating on technology supported science investigations (e.g. in the Personal Inquiry project).

The review presented here will describe and assess findings from this work, and review the methods employed in these studies. Methods of data collection adopted were aimed at generating rich descriptions of the interactions between learners and computers and include the use of video records and content analysis of discussion protocols. A number of analysis frameworks were employed in this work. Those reconsidered here include the Context, Interaction and Outcomes (CIAO) evaluation framework (see Scanlon et al. 1998a) and video and transcript analysis incorporating technical tools such as Transana and the application of Activity Theory and other socio-cultural approaches to the analysis of data collected while investigating complex settings. For each case included in this review this discussion includes some data presented illustrating how the method is used, a detail of the methods used for documenting and analyzing interactions, and a discussion of the theoretical underpinnings of the methods used. The chapter will conclude with a discussion of the implications for this work for the challenges we have in understanding learning (processes and outcomes).

E. Scanlon (✉)
Institute of Educational Technology, Open University, Milton Keynes,
Bucks, MK 7 6AA, UK
e-mail: e.scanlon@open.ac.uk

S. Puntambekar et al. (eds.), *Analyzing Interactions in CSCL*, Computer-Supported Collaborative Learning Series, DOI 10.1007/978-1-4419-7710-6_15,

15.1 Introduction

This chapter discusses the implications of a 20 year series of experiments at the Open University on the development of appropriate methods of studying computer supported collaborative learning and its evaluation. Computer supported collaborative learning (CSCL) has been the subject of study for more than 20 years. O'Malley (1995) produced the first edited collection of research studies after a groundbreaking meeting held with significant researchers in Maratea in Italy in the late eighties. Another significant milestone in the development of CSCL research was the consideration given to the subject in the nineties by the European Science Foundation sponsored programme on learning in humans and machines and published in Dillenbourg (1999) and Dillenbourg et al. (1996). The series of CSCL conferences meanwhile started in 1995. The development of the field of computer supported collaborative learning over the past 20 years has been driven by two factors: a growing understanding of the benefits of collaborative learning and the development of the communication capabilities of computers. Light and Littleton (1999) and Crook (1994) discuss the ways in which using technology impacts the learning situation in collaborative settings. Like Hmelo-Silver (2003) and Hmelo-Silver et al. (2008) we appreciate that new learning environments require an innovative approach to explore the ways in which they can enhance computer supported collaborative learning.

15.2 General Trends in Methods

The interest in computer supported collaborative learning coincided with the sociocultural turn in education over the same period i.e. the attention paid to the role of others in the learning of individuals which has been given extra prominence in theoretical accounts of learning. Recently interest in theories including situated cognition, constructivism and socio-cultural psychology have accompanied this shift. In addition people have found inspiration in Activity Theory as an approach to studying computer supported collaborative learning. These theoretical trends will be revisited later. First I will consider the use of different research methods in the study of CSCL.

For some time an emphasis on rigorously controlled experimental studies of computer-supported collaborative learning was evident (see O'Malley et al., 1996 for a review). In particular, many approaches arrived at trying to identify which features of computer supported collaborative learning produced high pre- to post-test shifts focused on whether there has been learning rather than how this learning occurred. In many studies, even where such shifts were identified, it was difficult to work out what were the significant features of the interactions with technology and/or other people contributing to the innovation's success. As part of our research programme we first identified an approach to evaluating the use of educational software which became useful in such situations. The approach which was developed required an in depth consideration of the features of the learning setting. We focused on using Context, Interaction, Attitudes and Outcomes (CIAO) as part of an evaluation framework (see Scanlon et al. 1998a). The approach taken in this framework was to consider and collect information from a variety of sources both qualitative and quantitative. The dimensions were context (rationale for use), interaction (detailed protocols including utterances and actions), attitudes and outcomes (measures of both cognitive and affective aspects of the result of using the program). Part of this approach, in terms of the evaluation of use of technology, was a developed understanding of the important role played by learners' perceptions of a setting. In a number of settings we established that learners' perceptions of collaborative learning situations were a significant factor (see e.g. Scanlon et al. 1992, 1998b).

Research in this area has a variety of purposes: investigating the benefits of computer supported collaborative learning, determining the mechanisms of collaborative learning, gaining evidence to improve software design or the provision of educational guidelines and the design of good tasks for collaborative learning. Different communities are involved in the conduct of this research ranging from computer scientists involved in human-computer interaction research (HCI), psychologists studying the mechanisms of thought and learning, and educationalists trying to design effective collaborative learning settings.

An experimental approach provides important insights but we need to do more to reveal the interactions that emerge between students, instructors, tasks and tools. As noted above there have been many experimental studies that may help us to understand the potential benefits of CSCL and to determine the mechanisms of collaborative learning (see e.g. O'Malley 1995; O'Malley et al. 1996). However, they are of limited use in terms of software design and educational guidelines for the development of CSCL software or activities.

The implication of these considerations is to encourage the development of a multi-faceted approach to investigating computer supported collaborative learning. It involves investigating collaborative learning from a range of perspectives: the learners, the teacher or instructor and the researchers. Where possible our studies of computer supported collaborative learning have extended to consideration of naturalistic settings, although sometimes laboratory based studies are used when particular variables need to be controlled or very detailed collaboration records made using specialized recording equipment.

This approach acknowledges the importance and influence of the context on the ways in which learners collaborate and that learners and teachers have different perceptions, understandings and expectations of learning situations. All these features may affect the ways in which learners behave during computer supported collaborative interactions. In order to understand the nature of the interaction we need to understand these differences. The learning environment can be thought of as consisting of several contexts which mediate individual learning and peer interactions in a variety of ways. In addition we have found that, rather than considering only the outcomes of a learning experience or a snapshot of the activity it is desirable where possible to develop a detailed picture of how individuals in a group situation interact and how those interactions develop over time.

The research methods we have adopted include videoing, interviewing and the use of questionnaires and tests of outcomes both cognitive and affective. Most of the situations we have applied these methods to are those in which science is learned collaboratively (see e.g. Issroff et al. 1997; Scanlon et al. 1996; Scanlon et al. 1998b; Blake et al. 2003; Scanlon et al. 2004) but we have also considered other subject areas (McAlister et al. 2004; Waycott 2004).

15.3 Case Studies

The cases reviewed in this chapter include a set of experiments on the technological mediation of collaborative learning. The components of the mediation include access to shared simulations and a variety of types of access to communication with co-learners. These experiments required the development of an approach to the study of computer supported collaborative learning and the development of an appropriate evaluation methodology for use in studying these complex settings. As noted above, studying computer supported collaboration requires a consideration of both process and outcomes. We have conducted a variety of naturalistic and lab based studies of collaboration.

It is our experience that introducing technology into a setting can have both predictable and unpredictable effects. Properties of the technology such as communication or simulation capacity can be used to plan a problem-solving activity where such properties can offer some predicted advantages. However using technologically mediated collaboration changes the nature of the activity in ways we can not predict. To study this in the past we have conducted work on this theme looking at situations where members of a problem-solving group are physically separated then reconnected via combinations of computer and communications technology to work collaboratively on the simulation. The particular focus in this work was on trying to understand how students could use a system which allowed them to conduct variable based practical experiments in order to help them develop their knowledge and understanding of conceptually difficult topics. We have conducted a number of such studies of synchronous collaboration between adults working together on shared simulations. We wished to develop an understanding of the virtual space created by shared simulations and video communication tools for supporting collaborative work between people at a distance. The adults are physically separated and then reconnected by combinations of computer and communication technology. Many of these communication technologies and computer combinations used systems designed by Randall Smith such as Shared ARK (ShARK) (Smith 1992; Smith et al. 1991) and Kansas (Scanlon et al. 2005a). The Kansas system had similar functionality to the earlier ShARK system but in the former the audio and video are integrated with the interface while in the case of ShARK the audio/video was output to a device separate to the computer screen. These were prototype technologies for allowing students to work together at a distance from each other via a shared simulation while maintaining voice and eye contact. (Other systems such as co-Lab (van Joolingen et al. 2005) have similar properties but only allow text chat as the communication channel). Two series of studies were conducted- one using a physics simulation (Running in the Rain) and one on a statistics based simulation (Gameshow). Results were reported in Smith et al. 1991 and Scanlon et al. (1993) for the first simulation and Joiner et al. (2002) and Scanlon et al. (2005a) for the second. These case studies helped us explore the particular ways in which workspaces can be designed to maximize the beneficial effects of collaborative problem solving see Scanlon et al. 2001.

This approach depended on the collection of video data to allow the analysis of key features of problem-solving behavior within groups of students working on collaborative learning tasks. Our work in this area has been supported by developing data capture facilities for the group (the data capture suite, see Blake and Scanlon 2003) which support the collection of detailed records of interaction. The design of these facilities was inspired by our earlier collaboration on experiments at Xerox Europarc (see Buxton and Moran 1990). The studies used in this chapter to illustrate this approach include those of a computer-supported learning environment where we use video records of video-mediated collaboration. Rich data can be collected using video recording in this way and analyzed to increase understanding of computer-supported collaboration (see e.g. Lesh and Lehrer (2000) and Heath and Hindmarsh (2002) for further information on analysis methods). Although the focus of this chapter is on synchronous collaboration in the group we have also conducted work on near synchronous and asynchronous systems (see e.g. Holliman and Scanlon 2006).

15.4 Methods of Data Collection

Data collected included pre- and post-test questionnaires, video records and the detailed analysis of video material. The data capture suite was used to produce video records of the interaction. We captured video records of interaction between subjects together with a synchronous record of their computer screen. Video cameras recorded talk, task performance and records of collaboration. For example in our experiments with pairs of subjects working together each user's screen was captured together with each individual user's record of collaboration on video. For analysis purposes these different recordings were combined in a four-way matrix screenshot. The combined four way screen included video and has each participant's face (i.e. allows the tracking of their video conferencing record) and their shared computer screen. This made possible the simultaneous viewing of participants' verbal and non-verbal communication and their interaction with the computer simulation which allowed the detailed analysis of the sessions.

We were considering interaction between the subjects and their performance on the task. We transcribed the verbal protocols. We then considered each verbal protocol and related utterances made by participants to events while working with the computer simulation and any non-verbal communication such as eye contact. Eye contact events and other non-verbal communication were identified by reviewing video records. Utterances on the video were then transcribed and utterances related to the events working with the simulation or working with partners captured on participants screens. We devised a categorization of the utterances according to the type of activity the participants were engaged in according to whether they involved the interface, the task or social interaction. Within these categories each utterance was further classified as meta-level activity, specific activity and recovery. So, for example, the activity of generating a hypothesis would be classified as meta-level while recovery might be finding your way round any mistaken view of an interface and specific might be talk about some aspect of the task. These utterances were then further assigned to the categories planning, running experiments and other.

Various aspects of the use of the ShARK system for problem solving are described in Taylor et al. (1991) and in particular some particular features of the dialogue and the Kansas system in Scanlon et al. (2005a).

15.5 Analysis

The classification activity required protocols of the talk and many repeated viewings of the video and therefore was very time consuming. This hand coding approach was taken for the ShARK experiments. This approach did reveal some interesting features of the unique augmented space for collaboration and problem solving which had been created, with its mixture of overlapping (shared) and non overlapping work areas. The use of a video channel was expected to help with encouraging non-interface-specific activity, and that proved to be the case as we were able to demonstrate that meta-level discourse about the task was accompanied by a higher level of eye contact than specific talk about the interface over the video channel. This is described further in the next section.

In addition during the analysis of the KANSAS experiments we used a video coding system entitled Transana to time-code transcripts and videos to perform a detailed analysis of activities taking place during the experiments as described below. Transana is a video transcription and analysis tool which provides a way to view video, create a transcript, and link places in the transcript to frames in the video. Tools are provided for identifying and organizing analytically interesting portions of videos as video clips, and for attaching keywords to them. Database and file manipulation tools that facilitate the organization and storage of large collections of digitized video are included. The interface has four windows. The video window allows the researcher to play, pause, and stop the video, and provides a shuttle bar for navigation within the video.

The structure of Transana is described in Scanlon et al. (2005a) as follows (Fig. 15.1):

> The transcript window is where one can view or edit the transcripts. Time codes can be added which link a frame in the video to a spot in the transcript. Once this is done, the highlight in the transcript will move as the video is played, highlighting the appropriate section of the transcript. The sound window displays a waveform, or a bitmap representation of the audio track for a given piece of video and it is very useful in inserting time codes. The database window's function is to provide an easy way to organize and manipulate the data. It is arranged in a tree structure. (Scanlon et al. 2005a, p. 172)

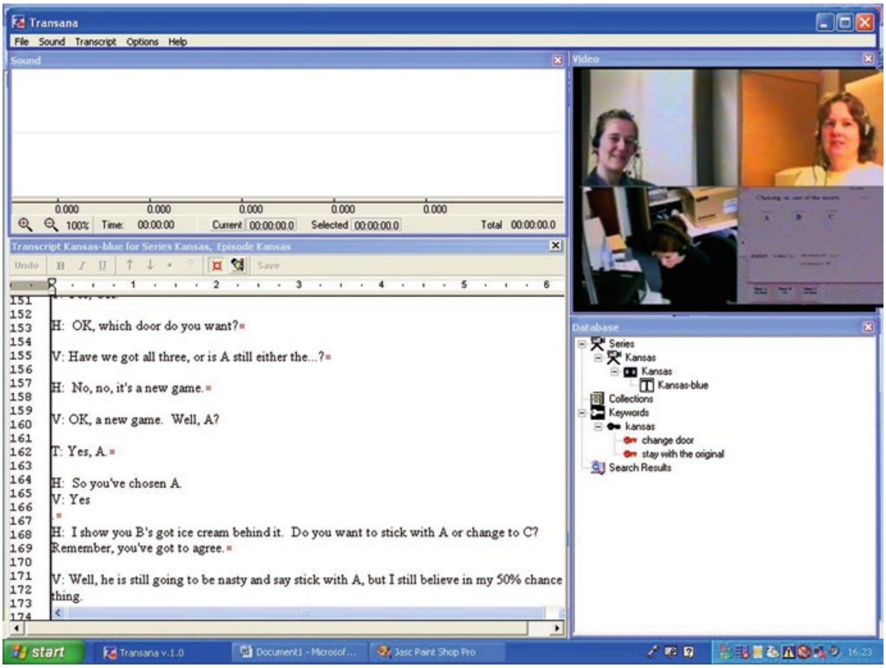

Fig. 15.1 Transana screen (reproduced with permission from Scanlon et al. 2005a)

15.5.1 Case 1: Running in the Rain

The motivating question for the research described in case studies 1 and 2 is 'what is different when members of a problem solving group are physically separated then reconnected via this type of computer and communications technology?'

The problem chosen for study in the Running in the Rain experiment was whether it was worth running in the rain or not. The research question being addressed was to identify the factors which were important in facilitating collaborative problem solving with this technology. We did this by comparing remote electronically mediated communication during collaborative problem solving, with that occurring during physical co-presence. The experiments involved 11 pairs of adults working at a simulation of this under-specified problem. In these experiments two users are in separate rooms with a workstation each, and communicate through a high fidelity, hands free audio link and with a camera/monitor device called a video-tunnel which enables both voice and eye contact through the use of a beam splitter and a mirror. An eight to ten minute introduction to the interface and the task was provided. The subjects (pairs of adults with some training in science) were then asked to make a joint decision of about when it is worth running in the rain and given a simulation containing a rain runner a rain cloud and a device to control the speed and direction of the rain and a device to measure the wetness of the runner. They could also switch the rain off or on and make the runner wider or narrower and

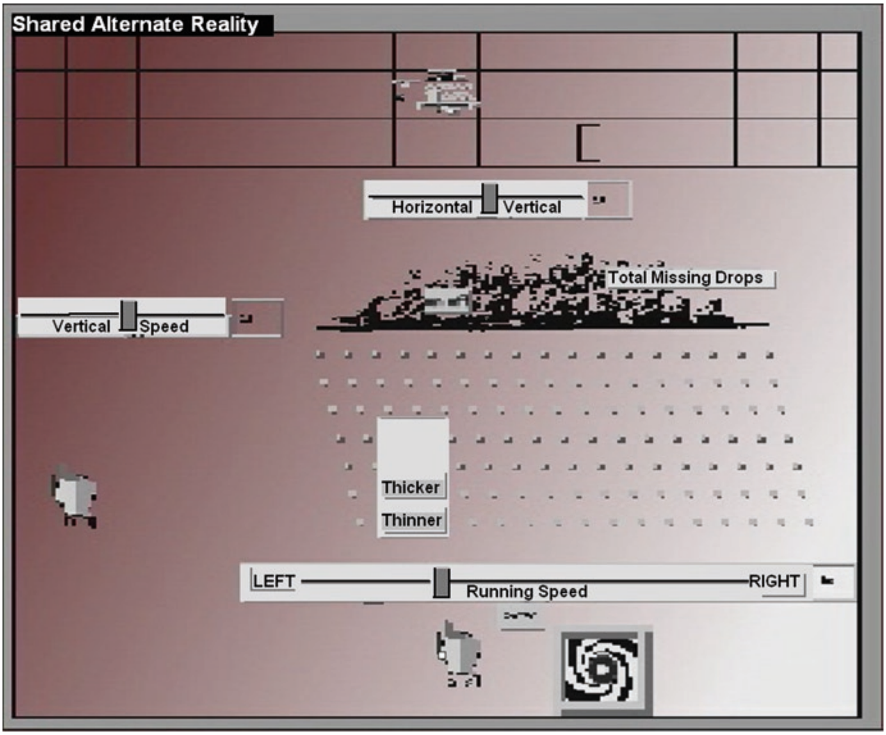

Fig. 15.2 A screen shot from the Running in the rain ShARK simulation

make the runner move. Participants were invited to use the simulation to test their ideas as they worked on a solution to the problem. They were then given 90 min to run experiments (Fig. 15.2).

All the subjects made progress with the solution. Scanlon et al. (1993) compared the problem solutions produced by pairs with the simulation and communication channels to those who produced a solution working only with paper and pencil. However, in addition to these demonstrated cognitive learning outcomes a main result was the importance of eye contact through video (a video-tunnel) in establishing successful collaboration. The landscape of collaboration created is that of an interesting shared space created by the technology. We referred to this as a kind of enhanced proximity because for the collaborators there was the possibility of being face to face by video while being side to side in looking at the shared simulation screen.

15.5.2 Context

In these first sets of studies the focus of our interest was on the different landscapes for computer supported collaborative learning created by these new technology

Table 15.1 Extract from protocol from post session interview

S1:	All the other factors we were talking about would be more important
R:	What's an example of a more important factor?
S1:	Well like how bothered they are about walking: whether they are about to have a heart attack; whether there's any shelter nearby…
R:	How far they have to go?
S1:	If they'd feel real embarrassed that their colleagues saw them running yes
S2:	If they were wearing mascara that runs they would

infrastructures. Even in these studies it was clear that the context of the problems chosen for examination with these systems played a considerable role in the way the subjects interpreted their tasks.

For the Running in the Rain experiment even in this laboratory setting there was extensive use of humor. See for example the interchange with reference made to the rain runner as 'Gene Kelly' singing in the rain, Glasgow a notoriously wet city in Scotland as the imagined location. Also

> 'One pair developed an extended fantasy of how they approached their role and task division with the male as hunter gatherer out chasing escaped features of the interface while the female collected harvests of raindrops' (See Scanlon et al. 1993, p 22).

The following exchange reproduced in Table 15.1 illustrates too some disconnection between the understanding of the mathematics necessary to establish a solution to the problem and the contextual factors that were felt still to be important to contribute to the problem solution. The Table includes an extract from a post-test interview between a pair of subjects with a researcher who asked them to consider their problem solution where they successfully drawn some conclusions about when it was worth running in the rain based on the experiments they had conducted. However, when he probed further other features of the simulation seemed to be more important to the subjects than the mathematical results they had established. Table 15.1 below shows an extract from their interview with the researcher (R) with comments from each member of the pair (S1, S2).

15.5.3 Case 2: Gameshow

The statistics problem chosen for exploration in this series of experiments was based on a well known problem the Monty Hall dilemma from the game show Lets make a deal! (Hoffman 1998). Subjects were told

> 'You are a game show contestant and your final challenge is to choose one of three doors. Behind one but only one of the doors is a Mercedes. You announce your selection but before you open the door the game show host helpfully opens one of the doors which was not the one you have chosen. It doesn't have a car behind it. Game show host gives you another chance. What should you do? Stick to your original choice or change?'

The pairs again communicated over an audio and video link and explored the problem with a shared simulated game show setting, a shared note-taking tool and a remote human host-the game show host. They were asked individually to make a prediction and to give a reason for that prediction and then given some time working with partners on a shared simulation relevant to this problem. They conducted experiments to help them agree on an answer to the problem. The consequences of each choice were displayed by the game show host. Similar data collection techniques and approaches to analysis were adopted as described above and which were used for the Running in the Rain set of experiments.

The study involved 24 pairs of subjects. They had completed at least a first degree and were volunteers from a pool of graduate students, researchers and educational technology developers. Two users were located in separate rooms with a workstation each and communicated through a variety of modes including either a high fidelity, hands free audio link only; video conferencing; or video conference with a possibility of allowing eye contact between the participants. In this way we were able to consider the effect of eye contact (Fig. 15.3).

Use of the video channel seemed to be associated with activity when subjects were not directly manipulating the interface. During some activities such as when data points were being collected, the video channel was not used at all. However it was used when the subjects were discussing what they observed or suggesting

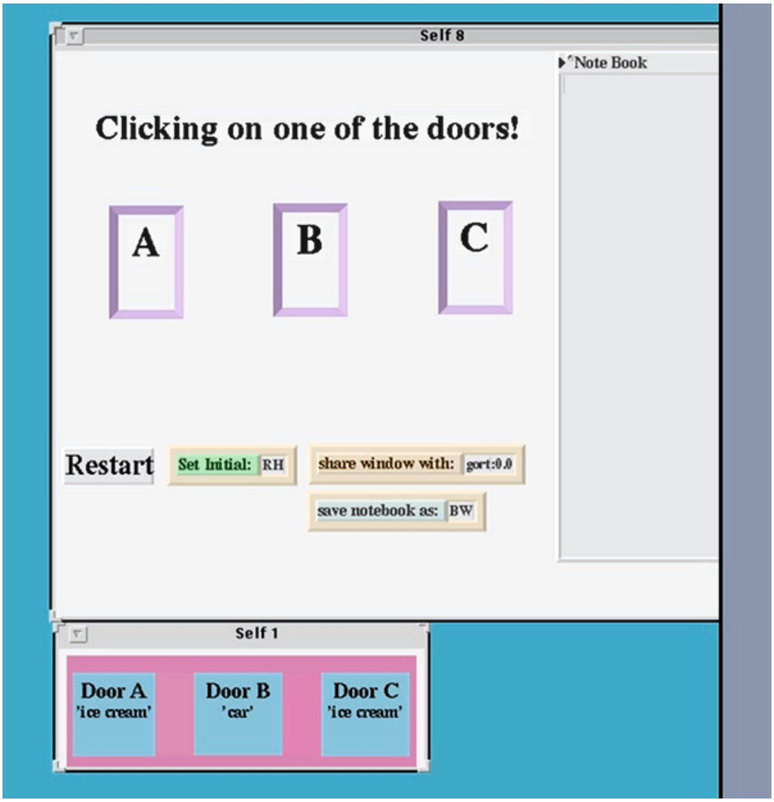

Fig. 15.3 A screenshot from the Gameshow Kansas simulation (From Scanlon et al. 2005)

hypotheses or planning experiments. In those circumstances subjects looked at their partner through the video. As in the running in the rain study we hypothesized that a video channel might encourage non-interface specific activity and were able to compare frequency of eye contact during of meta-level discourse about the task compared to periods when specific talk about the interface was taking place.

Again, all subjects made some progress with the problem and their understanding of the task. This is described in Scanlon et al. (2005). In this study we were able to consider the influence of mode of communication on successful problem solution and we were initially puzzled that subjects in our audio only condition was equally successful as those in the video tunnel condition with eye contact with less good outcomes from the video condition without eye contact we concluded that eye contact facilitating mutual gaze indeed was important, but that subjects were able to accommodate to the audio condition. Our findings were that some pairs using audio were fairly terse, and the use of audio did require more interchanges clarifying task division. This does not indicate that there is a particular problem with video which does not support eye contact, but shows that video with eye contact presents more opportunity for negotiating agreement

However, other factors appear to be important here too. Subjects often make explicit the idea that the more advantageous route of changing door choice entails the higher emotional risk of leaving a choice that turns out to have been right in the specific instance rather than following their statistical analysis and using the results of their experimental investigation. One pair had completed an investigation establishing that their chances of winning would be increased by changing their choice was interviewed after completing their investigation. Table 15.2 below shows an extract from their interview with the researcher (R) with comments from one of the pair.

When explored further this behavior turned out to be due to the perception that 'you'd feel so bad if you changed and then it won'. Another participant also commented 'Um if you watch game shows which I do quite a bit most people do stick to their original choice.'

The next section of the chapter will illustrate a shift in the type of studies conducted within the group to those where context and mobility have played an ever increasing role.

15.5.4 Discussion: Dealing with Complex Learning Settings

Our approach to making sense of the new spaces afforded for collaborative learning by technology rich settings has involved a multifaceted approach to experimental design, data collection and video analysis. We have illustrated how this approach gave us some insight into collaborative learning in a number of settings, in particular those involving complex problem solving. However in the exploration of these learning settings a number of considerations emerge. We agree with and Cakir and Stahl (2009) that

Table 15.2 Comments from one of a pair of subjects on completion of the Gameshow simulation

R:	It looks to me like you have more chance changing
S1:	Yes I wouldn't do that myself though
R:	Wouldn't you?
S1:	No you'd stick
S2:	I'd stick all the time

> While quantitative approaches can be effective in testing model-based hypotheses, they seem less appropriate both for exploring the problem of interactional organization and for investigating interactional methods p. 4

and with Suthers et al. (2007) that there is a need for a more eclectic approach to the study of how the work of collaboration gets done in rich technological and social environments. The approach Suthers (2006) proposes is a hybrid of experimental, descriptive and design methodologies. These concerns reflect a continuing debate since the late 1990s (see e.g. Littleton and Light 1998) on the best methods to be employed in analyzing productive interactions in computer supported collaborative learning.

The affordances offered by mobile technology have been our most recent challenge. The new settings for learning in which mobile technology is being used are a particularly complex area for analysis. These technologies create a new space for learning which is not attached to the constraints of time or place, for example, such spaces which appear when considering the use of mobile devices in an informal or semi-formal learning setting such as a museum or a field trip. In the case of the museum, the space for learning includes both the physical space of the museum and a virtual space which could include additional learning resources (e.g. multimedia additions to displays and the opportunities for collaborations with others either physically or virtually). There are challenges in constructing appropriate analytic frameworks for studying such settings recognizing the complexity of the concept of context (see Arvaja, this volume).

We needed to identify a framework which would allow us to consider such complex settings and to document the impact of the introduction of mobile technology to a collaborative learning situation. There are a number of challenges with this as we need to understand both how the complex setting functions to influence learning and the learners' perception of the purpose of the device and how it is used and adopted by them. Currently, our view is that Activity Theory approaches offer the best possibilities in this regard. Mwanza-Simwami (2009, p. 101) Suggests that 'activity theory seeks to explain the social and cultural embeddedness of human activities by linking them to issues relating to the motives of those involved in carrying out activities, and, the nature of the relationships that exist between and among those participating in the activities.'

We had experimented with this Activity Theory approach in a number of retrospective analyses of several case studies of our work including some case studies of computer supported collaborative learning in science (Issroff and Scanlon 2001, 2002). Building on the work of Vygotsky (Vygotsky 1978, 1987), Activity Theory allows us to consider learning involving a subject (the learner), an object (the task), and tool or mediating artifacts which mediates the subject and object. Engeström's Activity Systems (Engeström 1987) are a development of Activity Theory which offer a way of a representing the social context which influence human actions, as they are mediated by the rules of the community and division of labor within the community. It is useful as a way of unpacking the influence of the social and technological setting on learning (see e.g. Scanlon and Issroff 2005, Scanlon et al. 2005b).

15.5.5 Case 3: Mobile Collaborative Learning in Formal and Informal Learning Settings/Personal Inquiry

We are developing this Activity Theory informed approach further in our current work on school based learning of science using mobile devices. This investigation forms part of the Personal Inquiry (PI) project an ESRC/EPSRC funded research project being jointly conducted by teams at the Open University and the University of Nottingham, one of whose aims is to help 11–14 year old young people to use personal and mobile technologies to make their science learning more accessible and more effective (see Scanlon et al., in prep.). The Personal Inquiry project started in 2008 and since then we have supported over 200 students involved in a range of personal inquiries involving topics such as diet (which involved students in making predictions as to the nutritional quality of their diet, testing these by keeping a diary of their meals and snacks and then using this to calculate their nutritional intake), microclimates (involving students in deciding where in the school grounds would be the best locations for different types of activity, and then collecting scientific data at different locations in order to test their predictions) and Urban Heat Islands the subject of the inquiry described in this case study. This project was run over two iterations we have involved around 78 students for the first iteration and 57 students in the second iteration.

We are using Activity Theory frameworks to examine the interactions, contradictions and tensions that arise when pupils engage in inquiry learning using personal technologies in and out of school. The central relevant Activity Theory concept we use is tool mediation as expressed by Mwanza-Simwami (2009) 'Through the development and use of conceptual tools (such as human language and software applications) human beings externally they transform the activity and their own and others perception of the activities they are engaged in. At the same time by developing and using physical tools (such as PDAs mobile phones) human beings externally transform the activities they engage in (and…) studying human activities as developmental processes is crucial for identifying changes and activities that exist in an activity. Contradictions serve as the means by which new knowledge about the activity being examined emerges.' (p. 100)

We have produced a personal inquiry toolkit designed to enable learners to investigate personally-relevant questions, by gathering and sharing evidence, visualizing rich information, and engaging in informed debate. The toolkit incorporates structured collaborative activities adaptable by teachers to support the learner in working and making conceptual links across different activities (e.g. reading, data collection, visualization, and discussion), technologies (such as data probes, or web pages) and contexts (school, field work). Our research is considering questions such as how can scripted personal technologies including such toolkits be designed to support effective learning across transitions between formal and informal settings, and how are such technologies appropriated as tools for learning?

For example, we considered a case study of school students collaborating on conducting a Geography investigation using our location-based personal inquiry toolkit on a field trip where students were developing hypotheses as to how temperature varies across an urban area, making measurements and observations, and using these to explain any variations in temperature observed. The field trip centered on the exploration of the Urban Heat Island effect. The activity was part of the pupil's externally assessed coursework for their end of year examination. We have produced an account of how the introduction of technology into the activity changed and mediated the activity, and in addition we considered how the students integrated the toolkit into their performance of the activity. The introduction of mobile technology to the activity made different measurements and analyses possible. The introduction of technology into this new field trip activity changed the activity by making it possible to conduct a more extended investigations. The young people were able to compare their home city with another nearby urban area. They collected primary data and learned the skills needed to use a variety of different data sensors and the technology resourced the pooling and visualization of data.

We used a range of data collection methods to document and observe this collaborative learning event. We are able to draw on videotaped observations, the data students collected, and the notes and products created by the learners and teachers in the trials. Video records of workshops involving teachers, pupils and others during the design of our personal inquiry framework are also available as data.

(Like Reiser et al. 2000 we found the different agendas arising from the teachers' and university researchers' concerns involved in the study instructive.)

We have examined the ways in which the outcomes of the case study and the participatory design workshop have influenced our design decisions. We also drew on stakeholder interviews (with pupils and teachers) which focus on a reflection of the participatory design process.

One researcher reflecting on the process described it as follows

> What the process felt like to me was a series of meetings with individual stakeholders, that then intensified into repeat meetings with those who were likely to be responsible for taking the particular ideas forward. [...] from a series of meetings between the OU team and individual stakeholders, we then hit a more intensive period. The first sort of round of meetings exposed the areas of joint concern and suggested some ways forward. But then when we started trying to design things in earnest, the pace hotted up and there would be ideas that came up in one meeting that might be discussed back in the project team and that might be discussed again in a school meeting. Done some work on something and then come back with something that could be presented at the next meeting.

The new activity which was developed did have several good points, including the possibility of doing a richer comparison of data. However, this did in itself add complexity to the exercise. Using the Activity System approach we were able to focus on different communities' understanding of aspects of the activity (communities such as teachers and pupils.). Pupils had difficulty in operationalizing their understanding of what a hypothesis was based on their previous work in both science and geography. The teachers had a strong focus on assessment criteria which were set by the external examination board. This introduced certain constraints into the activity, in the area of collaboration. There was good collaboration between groups when they were on the field trips. However the tension within this particular piece of work was that the group element (joint working on the field trip) necessarily provided the foundations and main substance for the individual pieces of work pupils had to submit for their examination board coursework (Fig. 15.4).

In our use of Activity Theory we were influenced by the studies conducted by Waycott (2004) and Waycott et al. (2005) and in particular the approach taken to the study of mobility in informal learning settings. These studies report on how mobile technology was appropriated in a number of learning settings including an art gallery, reporting on 'how the two way nature of the process- how new users appropriate mobile technology and integrate it to suit their own purposes and how in turn those technologies change the way learners do things, shaping both their actions and their environment' (p. 12).

In this context, a small observational study of visitors to an Art Gallery exhibition on Landscape, Matter and the Environment, visitors used a Personal

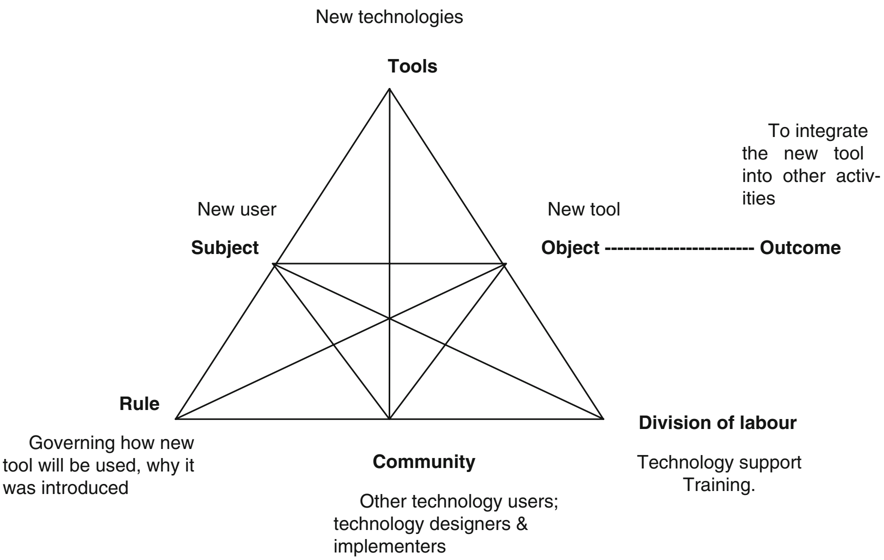

Fig. 15.4 An activity system representation of tool appropriation (Adapted from Fig 3.5 Waycott 2004)

Digital Assistant (PDA) containing background information in a variety of media about works on display, and also allowed for games and opinion polls to be used and offered the possibility of communication with others via standard text messages. In terms of the usability of the devices the PDA interfered with use of pen and paper for personal records and offered both possibilities and constraints to the activity of learning collaboratively in an Art Gallery. The use of multimedia content appeared to enhance the activity, expanding the type of information available to visitors. The text messaging option, which it was hoped would also introduce a new possibility to the activity 'did not successfully emulate the more dynamic and spontaneous verbal communication that visitors engaged in face to face. Also the technical difficulties, novel interface constrained the activity causing temporary breakdowns and shifts in focus from the activity to the tool itself.' (p. 13)

The emphasis placed by Activity Theory on the tool was very helpful in our analysis of the way activities are mediated by technology, as it shifted our attention away from interaction with the computer to the activity seen as a whole. We see great potential in the adopters of Waycott's (2004) approach to tool appropriation and integration to our analysis of the Personal Inquiry project by considering appropriation as an activity system. We intend to use the Activity System Tool Appropriation Model framework developed by Waycott (2004) and applied in Waycott et al. (2005) to help consider the constituent processes, resourcing and constraints introduced by the introduction of a new tool.

However there are challenges we face in other aspects of the interpretation of the data we are collecting in this situation that related to students learning trajectories. We need to construct an understanding of how our students develop (over an extended period of activity lasting a couple of months) their knowledge about urban heat islands and their understanding of how to conduct an inquiry. As a number of commentators have observed (see e.g. Mercer and Littleton 2007; Mercer 2008) knowledge is constructed over time and teachers use their talk to help bring together what might other be seen as a disconnected sequence of events, in our case a sequence of steps along the road to developing their answer to their personal inquiry questions. Our experience so far is that the Activity Theory approach is less suited to the exploration of such temporal sequences, although Greenhow and Belbas (2007) and Mwanza-Simwami (2009) express confidence in the possibility of applying activity theory concepts to complex settings and in particular to constantly challenging environments in which mobile learning occurs 'providing a mechanism for making the inter-relatedness of interaction processes more explicit' p117. Our current approach is to construct a series of trajectories recording over time how teachers and pupils made use of the different resources at their disposal including the representations of inquiry and the data visualization affordances made available by the use of the personal inquiry toolkit. Part of this involves, as Arvaja (this volume) describes, 'analyzing discourse temporally through extended dialogues, throughout the whole data, and identifying emerging bodies of knowledge or practices, in which are intertwined past and new knowledge.' This means that in our ongoing work we are taking a broader socio-cultural approach to the analysis of data in which the significance of resourcing and supporting the cumulation of understanding over time is the main aim.

In the case studies reported in this chapter, also we have used a variety of outcome measures, and in some cases examined the artifacts produced as a result of collaboration. In this study we were able also to examine the coursework produced as a result of collaborative working on the field trip. Like Paavola, and Hakkarainen (2005) we consider it particularly interesting to examine the knowledge creation metaphor for learning and we are extending our preliminary analysis of the field trip to consider the technologies used to support different types of mediation during the collaboration process.

15.6 Conclusions

There are several implications for this work for methods of understanding computer supported collaborative learning. Our approach initially involved studying complex problems and producing detailed accounts of how the landscape for collaborative learning is changed by the introduction of technology. In particular we have illustrated in cases 1 and 2 *Running in the Rain* and *Gameshow* our approach to researching and developing an understanding the virtual space created by the shared simulations and video communication tools we developed for supporting collaborative work between people at a distance.

We have accrued considerable experience involving the detailed analysis of the rich multimedia data and multiple methods approach to experimental design. In this chapter we have not discussed the challenges presented by managing the complex multi-modal records of interaction. Cox (2007) has a variety of suggestions to help with this. In addition, such tools as the Tatiana environment developed by Dyke et al. 2009 offer the possibility of assisting in this process by creating artifacts representing particular analyses.

The analysis which underlies the CIAO evaluation framework stresses how important it is when studying computer supported collaborative learning to take a broad view of the range of possible learning outcomes. We have adopted where possible as shown in cases 1–3 an approach to the measurement of learning outcomes which involves both cognitive and affective outcomes. Concerns about the difficulty of producing appropriate evaluations of CSCL systems are echoed in Jorrín-Abellán et al. (2009).

However in our more recent work with mobile technologies being enrolled to support collaborative learning situations, we have faced several more challenges which are presented to the researcher in CSCL. We have illustrated in this chapter (in case 3) the way that an Activity Theory approach offers the possibility of attending to the rich, social and technological settings which are thereby created. In this we are agree with Avouris et al. (2007) on the prospects of this approach.

In this latter case, however, we notice also a challenge to our approach to the measurement of outcomes, in particular a shift in what view of learning can be undertaken or supported in such settings. In the Art Gallery example of informal learning with adults using mobile devices mentioned within case study 3 it becomes even more difficult to develop appropriate methods to measure learning as it is difficult in such settings to predict in advance the goals of such informal visits. In the *Personal Inquiry* case, based as it is in a school setting we can have access to curriculum goals to help us design appropriate pre and post tests to measure learning outcomes but there remains the challenge of appropriate attribution of outcomes to particular sequences of interaction. Access to mobile devices allows the learning episodes to extend beyond formal classroom settings, through semi-formal fieldwork settings into interactions in the home.

In this way mobile and ubiquitous computing offer the prospect of yet more complex technology rich environments for collaborative learning. Mobile technologies, such as mobile phones and PDAs, have become ubiquitous and networked, and therefore offer distinctive opportunities for rich social interactions, context awareness and internet connectivity which have significance for learning. Some commentators predicted that 'learning will move more and more outside of the classroom and into the learners' environments, both real and virtual, becoming more situated, personal, collaborative and life-long' (Naismith et al. 2004, p. 5). As much of the research concerning the use of mobile technologies is: 'driven by an interest in the technical capabilities of new devices' (Naismith et al. 2004, p. 9) there is a need for detailed work examining the ways in which such technologies mediate and transform learning and teaching processes and practices. We see potential in mobile technology for supporting the transitions between the different sites (home, school and other) where collaborative learning may occur. It is in the tracking of such interactions that we see our future challenges in the analysis of productive computer supported collaborative learning.

Acknowledgments Thanks are due to the adults and young people involved in our studies of computer supported collaboration. Thanks are also due to colleagues involved in this work over the past 20 years including Tim O'Shea, Randall Smith, Josie Taylor, Richard Joiner, Canan Blake, Kim Issroff and Ann Jones and Jenny Waycott. Thanks are also due to current colleagues at the Open University and University of Nottingham on the Personal Inquiry project. Figures 15.1 and 15.3 are reproduced with permission from Scanlon et al. (2005) Technology Mediated Complex Problem Solving on a Statistics Task, Learning Media and Technology, 30(2) 165–183 http://www.informaworld.com

References

Arvaja, M. this volume. Analyzing the contextual nature of collaborative activity. In S. Puntambekar, G. Erkens, & C.E. Hmelo-Silver (Eds.), *Analyzing interactions in CSCL: Methods, approaches and issues* (pp. 25–46) Springer.

Avouris, N., Fiotakis, G., Kahrimanis, G., Margaritis, M., & Komis, V. (2007). Beyond logging of fingertip actions: Analysis of collaborative learning using multiple sources of data. *Journal of Interactive Learning Research JILR, 18*(2), 231–250.

Blake, C., Davies, C., Jones, A., Morris, E., & Scanlon, E. (2003). Evaluating complex digital resources. *Association for Learning Technology Journal, ALT-J, 11*(1), 4–17.

Blake, C., & Scanlon, E. (2003). Enriching accounts of computer supported collaboration by using video data. *Association for Learning Technology Journal, ALT-J, 11*(2), 5–13.

Buxton, W., & Moran, T. (1990). EuroPARC's integrated interactive intermedia facility (iiif): Early experiences, In S. Gibbs & A. A. Verrijn-Stuart (Eds.), *Multiuser interfaces and applications*, *Proceedings of the IFIP WG 8.4 conference on multi-user interfaces and applications* (pp. 11–34). Heraklion, Crete. Amsterdam: Elsevier Science Publishers B.V. (North-Holland).

Cakir, M., & Stahl, G. (2009). Interaction analysis of dual-interaction CSCL enrolments Proceedings of CSCL 09, volume 1.

Cox, R. (2007). Technology-enhanced research: educational ICT systems as research instruments. *Technology, Pedagogy and Education, 16*(3), 337–356.

Crook, C. (1994). *Computers and the collaborative experience of learning*. London: Routledge.

Dillenbourg, P. (Ed.). (1999). *Collaborative learning: Cognitive and computational approaches*. Oxford: Elsevier Science.

Dillenbourg, P., Baker, M., Blaye, A., & O'Malley, C. (1996). The evolution of research on computer supported collaborative learning. In P. Reimann & H. Spada (Eds.), *Learning in humans and machines,towards an interdisciplinary learning science*. Oxford: Elsevier Science.

Dyke, G., Lund, K., & Giradot, J. (2009).Tatiana: An environment to support the CSCL analysis process. In C. O'Malley, D. Suthers, P. Reimann, & A. Dimitracopoulou (Eds.), Computer supported collaborative learning practices, *Proceedings of CSCL 2009, Rhodes.*

Engeström, Y. (1987). *Learning by expanding: An activity-theoretical approach to developmental research*. Helsinki: Orienta-Konsultit.

Greenhow, C., & Belbas, B. (2007). Using activity-oriented design methods to study collaborative knowledge building in e-Learning courses. *International Journal of Computer-Supported Collaborative Learning, 2*, 363–391.

Heath, C., & Hindmarsh, J. (2002). Analysing interaction: Video ethnography and situated conduct. In T. May (Ed.), *Qualitative research in action*. London: Sage.

Hmelo-Silver, C. (2003). Analyzing collaborative knowledge construction: Multiple methods for integrated understanding. *Computers & Education, 41*(4), 397–420.

Hmelo-Silver, C., Chernobilsky, E., & Jordan, R. (2008). Understanding collaborative learning processes in new learning environments. *Instructional Science, 36*(5–6), 409–430.

Hoffman, P. (1998). *The man who loved only numbers: The story of Paul Erdos and the search for mathematical truth*. New York: Hyperion.

Holliman, R., & Scanlon, E. (2006). Investigating co-operation and collaboration in near synchronous computer conferences. *Computers & Education, 46*, 322–335.

Issroff, K., & Scanlon, E. (2001). Case studies revisited- what can activity theory offer? In P. Dillenbourg (Ed.), *Proceedings of First Euro-CSCL Conference*, Maastricht

Issroff, K., & Scanlon, E. (2002). Using technology in higher education: An activity theory perspective. *Journal of Computer Assisted Learning, 18*, 77–83.

Issroff, K., Scanlon, E., & Jones, A. (1997). Two empirical studies of computer-supported collaborative learning in science: Methodological and affective implications. In R. Hall, N. Miyake, & N. Enyedy (Eds.), *Proceedings of CSCL 97 The Second International Conference on Computer Support for Collaborative Learning* (pp. 117–123), Toronto, December 10–14.

Joiner, R., Scanlon, E., O'Shea, T., Smith, R,. & Blake, C. (2002). Evidence from a series of experiments on video-mediated collaboration: Does eye contact matter? In G. Stahl (Ed.), *Proceedings of CSCL 2002* (pp. 371–378), Boulder, Lawrence Erlbaum Associates.

Joolingen, W. V., deJong, T., & Manlove, S. (2005). Co-Lab: Research and development of an on line learning environment for collaborative science discovery learning environment. *Computers in Human Behavior, 21*(4), 671–688.

Jorrín-Abellán, M., Stake, R., & Martínez-Monés, A. (2009). The Needlework in evaluating a CSCL system: The Evaluand oriented Responsive Evaluation Model, In C. O'Malley, D. Suthers, P. Reiman & A. Dimitracopoulou (Eds.), Computer supported collaborative learning practices, *Proceedings of CSCL 2009, Rhodes*.

Lesh, R., & Lehrer, R. (2000). Iterative refinement cycles for videotape analysis: Conceptual changes. In R. Lesh & A. Kelly (Eds.), *Handbook of research design in Mathematics and science education*. New Jersey: Lawrence Erlbaum Associates.

Light, P., & Littleton, K. (1999). *Social processes in children's learning*. Cambridge: Cambridge University Press.

Littleton, K., & Light, P. (1998). *Learning with computers: Analysing productive interaction*. London: Routledge.

McAlister, S., Ravenscroft, A., & Scanlon, E. (2004). Combining interaction and context design to support collaborative argumentation in education. *Journal of Computer Assisted Learning: Special Issue: Developing dialogue for learning, 20*(3), 194–204.

Mercer, N. (2008). The seeds of time: Why classroom dialogue needs a temporal analysis. *The Journal of the Learning Sciences, 17*(1), 33–59.

Mercer, N., & Littleton, K. (2007). *Dialogue and the development of children's thinking: A socio-cultural approach*. London: Routledge.

Mwanza-Simwami, D. (2009). Using activity oriented design methods (ADOM) in researching mobile learning. In G. Vavoula, N. Pachler, & A. Kukulska-Hulme (Eds.), *Researching mobile learning*. London: Peter Lang.

Naismith, L., Lonsdale, P., Vavoula, G., & Sharples, M. (2004). *Literature review on mobile learning*, NESTA Futurelab Report No 11, Bristol.

O'Malley, C. E. (Ed.) (1995). *Computer Supported Collaborative Learning*. Berlin: Springer.

O'Malley, C., Langton, S., Anderson, A., Doherty-Sneddon, G., & Bruce, V. (1996). Comparison of face-to-face and video mediated interaction. *Interacting with Computers, 8*, 177–192.

Paavola, S., & Hakkarainen, K. (2005). The knowledge creation metaphor – An emergent epistemological approach to learning. *Science & Education, 14*, 535–557.

Reiser, B. J., Spillane, J. P., & Steinmuller, F., Sorsa, D., Carney, K., & Kyza, E. (2000). Investigating the mutual adaptation process in teachers' design of technology-infused curricula. In B. Fishman & S. O'Conner-Divelbiss (Eds.), *Proceedings of the Fourth International Conference of the Learning Sciences* (pp. 342–349), Erlbaum, Mahwah.

Scanlon, E., Blake, C., Joiner, R., & O'Shea, T. (2005a). Technologically mediated complex problem solving on a statistics task. *Learning, Media and Technology, 30*(2), 165–183.

Scanlon, E., Colwell, C., Cooper, M., & Di Paolo, T. (2004). Remote experiments, re-versioning and rethinking science learning. *Computers & Education, 43*(1–2), 153–163.

Scanlon, E., & Issroff, K. (2005b). Activity theory and higher education: Evaluating learning technologies. *Journal of Computers Assisted Learning, 20*(6), 430–439.

Scanlon, E., Issroff, K., & Murphy, P. (1998b). Collaborations in a primary classroom: Mediating science activities through new technology. In K. Littleton & P. Light (Eds.), *Learning with computers: Analysing productive interaction*. Routledge: London.

Scanlon, E., & Jones, A. (in preparation). *Investigating personal inquiry: An activity theory approach*.

Scanlon, E., Jones, A., Butcher, P., Greenberg, J., Ross, S., Taylor, J., et al. (1998a). Learning with computers: Experiences of evaluation. *Computers & Education, 30*(1/2), 9–14.

Scanlon, E., Jones, A., & Waycott, J. (2005b). Mobile technologies: Prospects for their use in learning in informal science settings. *Journal of Interactive Media in Education, 21*(5), 1–17.

Scanlon, E., Murphy, P., Hodgson, B., & Whitelegg, E. (1992). A case study approach to studying collaboration in primary science classrooms. In H. Foot, C. J. Howe, A. Anderson, A. Tolmie, & A. Warden (Eds.), *Group and Interactive Learning*. Southampton: Computational Mechanics Publications.

Scanlon, E., Murphy, P., Issroff, K,. with Hodgson, B., & Whitelegg, E. (1996). Collaboration in primary science classrooms: Learning about evaporation, In G. Cottrell (Ed.), *Proceedings of the Eighteenth Annual Conference of the Cognitive Science Society* (pp. 295–300), July 12–15, University of California, San Diego, Mahwah, NJ: Lawrence Erlbaum Associates.

Scanlon, E., O'Shea, T., Smith, R., & Joiner, R. (2001). Technological mediation of science: Statistics and synchronous collaboration: Ark, Shark and Kansas. In R. Joiner (Ed.), *Rethinking collaborative learning*. London: Free Associates.

Scanlon, E., O'Shea, T., Smith, R. B., & Li, Y. (1997). Supporting the Distributed Synchronous Learning of Probability: learning from an experiment. In R. Hall, N. Miyake, & N. Enyedy

(Eds.), *Proceedings of CSCL97 The Second International Conference on Computer Support for Collaborative Learning* (pp. 224–230), Toronto, December 10–14.
Scanlon, E., O'Shea, T., Smith, R., O'Malley, C., & Taylor, J. (1993). Running in the rain-can a shared simulation help to decide? *Physics Education, 28*, 107–113.
Smith, R. (1992). A prototype futuristic technology for distance education. In T. O'Shea & E. Scanlon (Eds.), *New developments in educational technology*. Berlin: Springer.
Smith, R., O'Shea, T., O'Malley, C., Scanlon, E., & Taylor, J. (1991). Preliminary experiments with a distributed, multi-media problem solving environment. In J. M. Bowers & S. D. Benford (Eds.), *Studies in computer supported cooperative work: theory, practice and design*. Amsterdam: Elsevier.
Suthers, D. (2006). Technology affordances for intersubjective meaning making: A research agenda for CSCL. *International Journal of Computer Supported Collaboration, 1*(3), 315–337.
Suthers, D., Dwyer, N., Medina, R., & Vatrapu, R. (2007). A framework for eclectic analysis of collaborative interaction. In C. Chinn, G. Erkens, & S. Puntambekar (Eds.), *The computer supported collaborative learning (CSCL) conference 2007* (pp. 694–703). New Brunswick: International Society of the Learning Sciences.
Taylor, J., O'Shea, T., O'Malley, C., Scanlon, E., & Smith, R. (1991). Discourse and harmony: Preliminary findings in a case study of multi-media problem-solving. *Paper presented at the Conference on Problems of Support, Survival and Culture*, Amsterdam, April, 15–36.
Vygotsky, L. S. (1978). *Mind in society: The development of psychological processes*. Cambridge: Harvard University Press.
Vygotsky, L. S. (1987). *The collected works of LS Vygotsky*. New York: Plenum.
Waycott, J. (2004). *The appropriation of PDAs as learning and workplace tools*,' Open University, unpublished PhD thesis.
Waycott, J., Jones, A., & Scanlon, E. (2005). PDAs as lifelong learning tools: An activity theory based analysis. *Learning, Media and Technology, 30*(2), 107–130.

Chapter 16
Tracing Interaction in Distributed Collaborative Learning

Daniel Suthers and Richard Medina

Abstract In order to understand how entanglements of the activities of multiple individuals in technology-mediated environments result in learning, it is necessary to trace out activity that may be distributed across time, space and media. Multiple analytic challenges are encountered, including the distributed nature of the data, the contingent nature of human behavior, understanding nonverbal behavior, selective attention to large data sets, and multi-scale phenomena. This paper offers approach to analysis that was developed in our laboratory to address some of these challenges. In order to unify multiple data sources into one analytic artifact, we found it useful to abstract from media-specific units of analysis (e.g., adjacency pair, reply) and represent our data using "contingency graphs" that capture the potential ways in which one act can be contingent upon another. Contingency graphs serve as abstract transcripts that record distributed interaction in one representation. This chapter describes the contingency graph representation, gives an example of its use in analyzing the development of shared representational practices, and discusses further challenges. Important questions remain concerning the extent to which interactional accounts can remain productive as we grapple with larger data sets and emergent phenomena, and whether a productive interplay between interactional and aggregate accounts are possible that together inform design.

16.1 Introduction

This chapter describes a framework for analysis that we developed to support research on representational affordances for collaborative learning. This research has generally been concerned with a fundamental two-sided question: how are the affordances of designed media appropriated for intersubjective meaning-making, and

D. Suthers (✉)
Laboratory for Interactive Learning Technologies, Department of Information and Computer Sciences, University of Hawai'i, Manoa, HI, USA
e-mail: suthers@hawaii.edu

S. Puntambekar et al. (eds.), *Analyzing Interactions in CSCL*, Computer-Supported Collaborative Learning Series, DOI 10.1007/978-1-4419-7710-6_16,

how can the availability and salience of these affordances influence meaning-making processes? This generalized research question has various instantiations in the settings that we study. The research program began with a series of laboratory studies of "representational guidance" in dyadic problem solving and learning that showed the influence of notations on certain categories of behavior (Suthers and Hundhausen 2003). These studies were framed within an experimental paradigm that manipulated notational systems to observe effects on quantitative measures of interaction, but did not analyze how the participants collaborated with these notations. We therefore undertook analyses of interaction to understand issues such as how argumentation and problem solving can take place through joint synchronous manipulation of shared workspaces (Suthers 2006a), how participants in an asynchronous setting not only share information but come to agreement on its interpretation (Suthers et al. 2007b), and how representational practices are invented and develop in such environments (Dwyer and Suthers 2006; Medina and Suthers 2008.

During the same time, our laboratory was responsible for developing and supporting a statewide community of public school teachers through an online environment that included collaborative workspaces and shared resources (hnlc.org; see Suthers et al. 2004; Suthers et al. 2007), and we have applied a similar environment to a community-oriented approach to online graduate education (discourse.ics.hawaii.edu). In both of these efforts, we sought to embed task-oriented groups (e.g., school teams and courses, respectively, as well as special interest groups) in common digital environments that support opportunistic formation of social relationships and resource sharing. The objective is to enable the formation of "transcendent communities" beyond the scope of the teams or courses that brought members to the environment in the first place (Joseph et al. 2007). This objective led to our interest in analyzing boundary spanning and other phenomena concerning how new social relationships form in online environments and their consequences (Suthers et al. 2009).

Although diverse in terms of the settings and questions addressed, these studies share similar analytic challenges, some of which are addressed in this paper. In each case it is necessary to find the phenomena of interest in the entangled trajectories of the activities of multiple individuals—activity that is distributed across time, space and media. Our framework is based on several simple ideas: that it is useful to gather together the various traces of activity that might otherwise be distributed across data sources into one analytic artifact; that this artifact should make evidence for interaction explicit while also indexing back to the original data sources; that this analytic artifact can support moving between levels of analysis; and that it potentially can serve as a common framework for dialogue between multiple analytic approaches. This chapter begins by acknowledging some theoretical assumptions concerning our analytic program. It then covers the nature of data required to apply the framework, the analytic challenges we are confronted with, and how the framework begins to address these challenges. An example of multi-level analysis of dyadic problem solving is given, this being the application in which our approach is more completely worked out. The chapter concludes with discussion of the challenges that remain, and a broader view of the potential value of the framework.

16.2 Theoretical Assumptions

This research is concerned with *collaborative and networked learning* in *technology-mediated environments*. Below we discuss some of the assumptions behind the research program, as they have influenced the approach taken.

We are generally interested in learning "in the context of joint activity" (Koschmann 2002). This learning may occur at various social granularities and types of "joint activity," ranging from small numbers of tightly coupled collaborators who maintain a "joint conception of a problem" (Dillenbourg 1999; Teasley and Roschelle 1993 or engage in "group cognition" (Stahl 2006), to learning in "virtual communities" (Renninger and Shumar 2002) and in loosely associated networks of individuals (Castells 2001; Jones et al. 2006). There are several ways in which social contexts can be seen as contributing to learning, briefly reviewed in (Suthers 2006b). We are particularly interested in processes in which individual attempts at meaning-making influence others through the technological environment and lead to intersubjective meaning-making. To understand technology mediated collaborative and networked learning we need to trace out trajectories of intra- and inter-subjective meaning-making and relate these to the media used.

Our approach to understanding individual and collective trajectories of meaning-making is influenced by ethnomethodology's program of identifying the methods by which "members" produce recognizable accomplishments (Garfinkel 1967) such as learning (Koschmann et al. 2005), and is also influenced by actor-network theory's program of "reassembling the social" (Latour 2005) from a network of associations between human and technological actors. But our program departs from these sociological concerns with how social order is constructed in interaction, because our program is oriented towards the problem of *designing for learning*. For example, "ethnomethodological indifference" and "relevance" (Koschmann et al. 2007) are tempered by our need to select events that are interesting from a learning perspective. While all episodes of interaction are potentially interesting to an ethnomethodologist as instances of social organization (hence their "indifference" as to the social setting studied), research in learning science will find episodes in which learning appears to take place to be more interesting, or at least interesting in a different way, than those in which learning is not apparent. While ethnomethodological accounts seek only to be relevant to members' accounts of their actions, our accounts must be relevant to the practices of educators and researchers as well as those of persons participating in the episodes analyzed. The design orientation admits of etic and prescriptive analyses.

"Learning" is understood broadly as a judgment that some worthwhile transformation of individuals has taken place (Suthers 2006b). Our analyses work back from this recognition to trace out the interactional accomplishment of meaning-making that led to that learning. We speak of "meaning-making" because we maintain that participants are continuously trying to make sense of their experience at multiple levels (e.g., concerning the task, interpersonal relations, and normative behavior; Bronckart 1995). Participants are engaged in "doing" meaning-making; the question of learning is an evaluation of the consequences of that doing.

The most basic units of data are acts by which participants *coordinate* between personal and public realms, including with each other. Other literature uses the term "contribution," but we desire a term that does not imply a conversational setting or a particular kind of participatory intention, and that is not biased towards production as the only kind of relevant action. For example, when a participant reads a message the personal realm is brought into coordination with inscriptions in the message, and when the participant writes a message, inscriptions are created in the public realm that are coordinated with the personal realm. We are influenced by Hutchins' idea of "coordination of [not necessarily symbolic] information-bearing structures" (Hutchins 1995, p. 118) in his theory of distributed cognition. However, the analytic approach outlined in this chapter does not require assumptions about the nature of the "personal realm": it only requires the assumption that an actor's actions may be contingent upon their history. Therefore, to avoid unintended theoretical baggage, we now use the term *act*, or more generally *event* (to include computer-generated events) instead of "coordination."

The basic relational unit of interaction for any analysis of intersubjective meaning-making is the event of a participant taking up a prior act or trace of that act as having some relevance for ongoing participation. We call this basic relational unit "uptake" (Suthers 2006a; Suthers et al. 2007). Uptake includes but is more general than "transactivity" (Berkowitz and Gibbs 1979) in that uptake need not be directed towards a particular other actor, and indeed can occur in realms of participation in which the originating actor is not participating. Also, the concept of uptake applies to intrasubjective relationships between events as well as intersubjective ones, We must infer uptake from observable contingencies between actions, which requires a separate representation of this evidence. Therefore, our approach is built upon the concept of contingencies as detailed later in this paper.

The technological medium is not neutral. A given medium offers certain affordances (potentials for action in relation to the actor, following Gibson 1977), of which salient affordances are expected to be the most relevant (Norman 1999). Affordances are the means for participation in social realms of activity as well as physical ones. Affordances are not deterministic: actors may choose to appropriate affordances in particular ways to enable their participation (Hamilton and Feenberg 2005; Suthers 2006b). However, the availability and salience of affordances of designed artifacts influence these choices in ways that are of interest to us as researchers and designers.

16.3 Data Requirements

In general, any research concerned with how distributed interaction leads to learning requires process data concerning individual participants' manipulation of media and the availability of these manipulations to other participants, as well as some kind of "outcome" data that helps us select interactions that were fruitful from a learning perspective. These two requirements are discussed below, along with implications for data representations.

Ideally, the data would make uptake relationships between acts coordinations explicit, but it is not always this easy. Some data, such as reply relations in discourse media, make uptake explicit, but generally people take up each other's contributions or traces thereof in ways that are often not explicitly apparent or recorded in the medium of interaction. Participants might reuse someone's phrasing, re-express their ideas, begin to attend to an issue after someone else raised that issue, react emotionally to a statement, etc. Uptake can be manifest in nonverbal forms, such as manipulating representational objects previously created in a shared workspace by editing, organizing or connecting them. Human action is contingent upon its context in many subtle ways. Therefore the data should record sufficient information that enables us to identify various contingencies between media coordinations that may evidence the presence of uptake.

The myriad of contingencies in human behavior can be overwhelming, and many will not turn out to be relevant to a given analysis. An analysis must be selective. One strategy we take is goal directed: we identify particular outcomes of interest—such as participants coming to agreement on an explanation for a complex phenomenon—and then work backwards to provide an interactional account leading to this outcome. Consequently, the available data must provide some means of identifying outcomes of interest for this goal-directed analysis. In the context of educational research, we expect that the selection criterion will be a measure of or judgment concerning learning.

To date, we have not made much use of demographic, dispositional, or developmental data on participants. We have focused on what was available in the interaction being observed. Interaction projects back to birth and the prior interactions of others, but where such data is not available, dispositional descriptions are convenient summaries of the expected future effect of prior interactions. The lack of an explicit means for including such data may be seen as a shortcoming of our approach, but our approach does not exclude bringing in such considerations in interpreting interaction. For now, we choose to see what interaction can tell us. This is partly a simplifying research strategy, but is also influenced by arguments that tracing out networks of associations mediated through interactions is more appropriate for discovery of unanticipated relationships and provides more realistic explanatory accounts than attribute-based approaches (Blumer 1969; Latour 2005; Marin and Wellman 2010; Reimann 2009; Reimann et al. 2011).

The specific data required depends a great deal on the nature of the environment being studied. The nature of the data that comes from distributed environments is the first source of our analytic challenges, discussed next.

16.4 Analytic Challenges

Multiple analytic challenges are encountered when attempting to expose interaction in distributed collaborative learning settings. (Here we focus on those challenges that are consequent of or generally present in such settings, and do not attempt to catalog all challenges specific to particular research programs.) The analytic challenges begin with the nature of the data itself: how do we make interaction apparent from the myriad of contingencies between acts that are distributed across time, space and media? Then, because the data may be multimodal we need to provide an account of nonverbal behavior. Our desire to scale up to larger data sets leads to questions of multi-scale phenomena. We discuss these challenges further below.

16.4.1 The Distributed Nature of the Data

Some analytic challenges come from the setting. We have chosen to study technology affordances for intersubjective meaning-making in settings that (a) have multiple notational resources for interaction and (b) are spatially distributed and may be quasi-synchronous or asynchronous. Property (a) means that we need to construe activity that may be recorded in multiple log files or formats as unified activity, because interaction is distributed across all mutable media (Suthers et al. 2003). We need a way to gather together data from various sources and derive an analytic artifact that enables us to "see" interaction more directly. Property (b) means that we cannot assume that the frame of reference is the same for everyone. We need to trace out activity with respect to individual frames of references and how they intertwine into phenomena at the group level.

16.4.2 The Contingent Nature of Human Behavior

Human action is contingent upon its "context" (physical environment, history of interaction, institutional and cultural-historical settings) in many subtle ways. For example, even in the constrained environment of a threaded discussion it is not sufficient to consider only the reply structure recorded in the media. A posting (message) can also be contingent upon other prior media coordinations. It can repeat lexical strings, re-use typographical conventions, or follow up on ideas of previous postings not limited to the one being replied to. A posting can also be related temporally to the timing of other postings and (in an aspect of interaction often neglected in the study of threaded discussions) can also be temporally contingent upon the *reading* of other messages (Suthers et al. 2007). The possible relationships get more complex in graphical workspaces, where for example the placement of a shape on the screen is contingent upon the prior placements of other shapes (Shipman and McCall 1994). Decisions must be taken concerning when the contingencies between media coordinations merit the appellation of "uptake" and therefore inclusion in the analysis of interaction.

16.4.3 The Meaning of Nonverbal Behavior

Other analytic challenges are related to the meaning of nonverbal behavior. In addition to writing statements or labeling objects in natural language, users of a multimedia environment can manipulate and organize representations in ways explicitly and implicitly supported by the environment, such as linking objects or placing them in spatial arrangements relative to each other. When are such manipulations merely "housekeeping" and when are they conceptually significant? Later in this chapter we provide an example of how we uncovered conceptually significant manipulations of a graph representation of evidence (see also Medina et al. 2009).

16.4.4 Selective Attention to Large Data Sets

A fourth analytic challenge is related to our objective of scaling up interaction analysis from single sessions of dyads and small groups to larger groups, longer time spans and multiple media. We are able to record a large quantity of interaction data, but to what should we attend in our analysis to make our work tractable? Some analytic methods include all of the data by aggregating events into frequency counts, yet risk obfuscating the specific trajectories of mediated action by which learning was accomplished in context. Sequential analysis reveals these trajectories, yet is usually done selectively, and selective analysis leads to questions of representativeness and coverage (but see Lee and Baskerville 2003 for arguments against basing generalization solely on sampling theory). How do we ensure that this necessary selective attention does not leave out important observations or phenomena?

16.4.5 Multi-scale Phenomena

Even in dyadic interaction, multiple scales of analysis are required. We can understand the moment to moment actions of each individual in the context of their environment, thread these actions into coherent accounts of individual trajectories of learning, and analyze interactions between trajectories that lead to meaning-making at the dyadic or small group level. As the time scale of such an analysis increases, we can observe the introduction or improvisation of new practices and their adoption and development by the group. As the number of persons involved in technology-mediated environments increase, phenomena that transcend the immediate interaction between individuals emerge, such as collective resources, practices and identity. The challenge is not only in understanding phenomena at a given temporal or social scale, but also in understanding relationships between phenomena across scales.

16.5 Analyzing Distributed Interaction with Contingency Graphs

Our approach to analysis addresses some of the above challenges through an abstract transcript representation that we call “contingency graphs.” This section describes contingency graphs, summarizes how we use them to address the analytic challenges, and gives an extended example.

16.5.1 Contingency Graphs

A contingency graph captures interaction in a medium-independent manner that can yet be annotated with information about media properties. This notation enables us to gather together activity that is distributed across media (and hence data sources) and across participants into one analytic artifact. A contingency graph consists of vertices (nodes) that represent events and directed arcs (links) that represent contingencies between these events. Multiple log files from multiple sources of data are merged into the contingency graph, retaining pointers back to the original data but enabling us to have a single abstract transcript that gathers together all potential interaction into one artifact that can be visualized and/or searched through computational processes. Although contingency graphs can be complex, they can be constructed on an as-needed basis, and can be visualized in different ways not limited to node-link diagrams.

16.5.1.1 Vertices: Events

Our approach is committed to an event-based rather than variable-based ontology (Reimann 2009; Reimann et al. 2011). The events represented by vertices may include any coordination with the medium enacted by participants, including (for example) the creation of media inscriptions (e.g., posting a message, making an object in a workspace), manipulation of those inscriptions (e.g., moving objects closer to each other), and perception of those inscriptions (e.g., opening a message to read it). The graph also may include computer-initiated events such as the display of inscriptions that come from other participants in an asynchronous environment, or events initiated by the technological infrastructure itself. Events record the actor, object acted on, and action taken, as well as temporal and other relevant information. Analysts can choose whether and how to treat computer-generated events differently.

16.5.1.2 Arcs: Contingencies

A contingency relationship holds when one or more events enable a subsequent event. The term “contingency” is chosen to indicate a sense of enablement in which human action draws upon but is not necessarily determined by elements of the

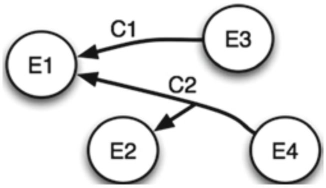

Fig. 16.1 Contingency relationships

environment, as discussed in section 0 and further below. Contingencies are represented in an acyclic directed hypergraph as hyperarcs (directed hyperedges) between events. Each arc points backward in time from a single origin to one or more destinations. For example, in Fig. 16.1, event E3 is contingent on event E1, and E4 is contingent on events E1 and E2.

Over the past 2 years, we have identified several types of contingency relationships between events (Suthers 2006a; Suthers et al. 2007; Suthers et al. 2007), including media dependencies, temporal and spatial proximity, representational similarity, and semantic overlap. These are discussed below.

Contingencies are most easily identified through similarities in events. The most straightforward approach is to construct contingency arcs between events that involve the same media entity. We call these *media dependencies*. For example, the events of opening and replying to a message are dependent on the event of creating a message, and the event of linking to or annotating a media entity depends on its prior existence.

In synchronous interaction, *temporal proximity* also implies relevance, such as in the typical reply structure of conversation (Sacks et al. 1974). People also exploit s*patial proximity* and *representational similarity* to manage interaction and express association (Dwyer and Suthers 2006; Shipman and McCall 1994). For example, if a representational element is given the same appearance as other elements (e.g., same color, location, or label), we construe this manipulation as contingent on previous uses of those visual attributes (e.g., adding an element to a group is contingent on the group's prior existence).

Tracing *semantic overlap* is more difficult. We can partially trace ideas by tracing the artifacts that express them, but actors may "transcribe" ideas to other artifacts, such as through quoting practices (Barcellini et al. 2005). More problematically for the analyst, ideas can be taken up and re-expressed in different ways. It is precisely these kinds of semantic transformations that are of greatest interest when studying (for example) the production of new knowledge in technology-mediated social networks.

16.5.1.3 Addressing the Challenges

Contingency graphs address the *distributed nature of the data* by collecting distributed data into a single analytic artifact that can then be inspected in the analysis of interaction. As a canonical representation, other benefits may accrue. Structural comparisons may be made across superficially different sources of data (e.g., different online communities). Later we will argue that contingency graphs also may serve as a boundary object for discourse between multiple analytic methods and disciplines concerned with interaction.

Clearly, contingency graphs help address the *contingent nature of human behavior* by providing an explicit representation of contingencies so that there is a basis for deciding which contingencies are relevant, using both automated tools and human judgment. Many approaches to coding discourse relations allow the analyst to simply assert the relation without requiring that the analyst specify the evidence on which this judgment was based. Explicit contingencies better support conversations about the choices made by analysts.

The challenge of understanding *the meaning of nonverbal behavior* has two aspects: selectional and hermeneutic. Contingency graphs can help us with the selectional problem of identifying nonverbal behaviors that merit closer examination to determine whether they are interactionally meaningful. As will be illustrated in the next subsection, some nonverbal behaviors of interest result in patterns in the contingency graph. Unlike separate log files, contingency graphs that unify all behaviors provide a uniform way to see patterns of nonverbal manipulations and situate them in the context of verbal behavior. (For a different visualization that identifies situated patterns of behavior, see Hmelo-Silver et al. this volume.) This turns out to be important because some patterns in nonverbal behavior are found in relation to verbal behavior. To address the second, hermeneutic problem one must go to the data in the original media formats and make interpretations. The contingency graph helps focus this effort, as will be illustrated shortly.

We address the problem of *selective attention to large data sets* by using goal-directed search in the contingency graph (e.g., tracing back from an interesting learning outcome), by using automated tools for finding relevant events and pathways in the graph, and through visualizations of the graph structure. Finally, we believe that contingency graphs will help address *multi-scale phenomena* through computational tools that enable analysts to find patterns at larger scales and relate them to aggregate phenomena. Our following example illustrates multi-scale analysis at the scales of episodes in a session down to micro-analysis. Extensions in the other direction, to be explored with our online community data, are still pending.

16.5.2 Example Analysis

This section describes a recent analysis using the contingency graph. It illustrates how we have begun to confront the challenges of working with records of distributed data and activity, finding significance in nonverbal actions, selectively attending to particular aspects of the data, and moving flexibly across multi-scale phenomena. The session we analyzed will be described, followed by a description of the log file and video data. This section ends with a detailed account of the analysis revealing a qualitative explanation of participants' convergence on a conclusion based on verbal and nonverbal interaction.

16.5.2.1 Source of Data

The case study presented here illustrates a pattern of interaction between two individuals engaged in a problem solving exercise while using a shared networked workspace environment (Fig. 16.2).

The individuals were participating in an experiment described in (Suthers et al. 2008). Using informational materials we provided in the workspace, the two participants (P1 and P2) worked to identify possible causes of a disease in Guam, ALS-PD (Amyotrophic Lateral Sclerosis-Parkinsonism Dementia complex). The session took place over the course of approximately 2 h. Participants were at different locations, and interacted in an environment that included a graphical evidence map and threaded discussion. Each participant's view of the shared environment was updated using a software protocol that enforced asynchronous interaction by distributing respective workspace changes at intermittent times during the interaction (participants were also able to manually request updates by selecting a refresh button). We undertook the analysis to account for ways in which participants both converged and diverged in their interpretations of causes of ALS-PD, by tracing out sequential patterns of representational practices enacted within the workspace. The analysis highlights an evolving transformation of a collaborative representational practice. These practices and the artifacts left in their

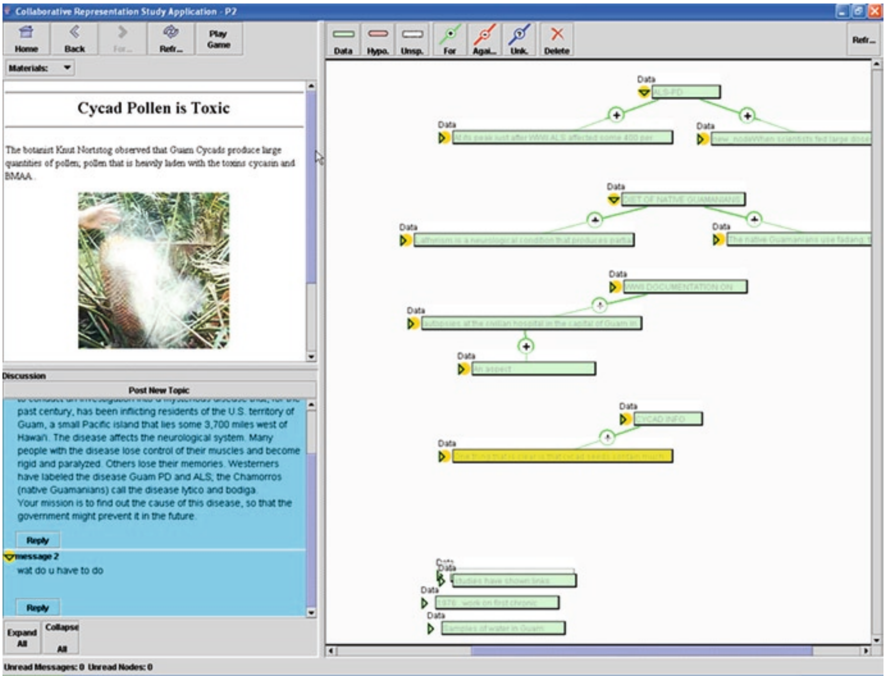

Fig. 16.2 Information source (*top left*), threaded discussion (*bottom left*) and evidence map (*right*)

wake provide an explanation for the conceptual convergence and divergence in the conclusions expressed by each participant.

16.5.2.2 Log File Description

The software used by these participants provided both threaded discussion and graphical evidence mapping tools. All actions performed by participants on and through the software media were logged. In addition to actions, network activity was logged to record when events recorded on each participant's machine were received and rendered on the other. With regard to the use of log files in this analysis, all acts have a corresponding event in the log file, can be attributed to a participant, can be identified in terms of the media object or objects implicated in the act, and are time-stamped relative to the machine on which the act was performed. Also, all network exchanges between participants' machines have a corresponding event in the log file. In addition to log files, we have a video record of each participants' screen, along with synchronized web-camera video of participants' faces, using Morae™ software.

16.5.2.3 Contingency Graph Construction

In this analysis, an initial contingency graph of *media dependencies* was automatically generated from log data by iteratively relating pairs of log events based on the following criteria:

- If two events share an object id (an artifact is edited, moved, etc.) then the later event is contingent upon the prior event.
- The event of linking two objects in the evidence map is contingent upon the most recent prior events available to the participant that modified the objects.
- The event of posting a discussion message as a reply is contingent on the event that created the replied-to message.
- If a discussion message contains a hyperlink to an evidence map object, then the message event is contingent upon the most recent event available to the participant that modified that map object.

To construct a contingency graph based on media dependencies we wrote a Java program that processes the log file using the above criteria, leveraging information stored in the log file such as the *object id*. Many other contingencies are available, but media dependencies are a convenient starting point for the analysis because they can be extracted automatically. Subsequent manipulations can add other kinds of contingencies. A visualization tool (Omnigraffle™) was used to display the contingency graph. See Fig. 16.3 for an example. See (Medina and Suthers 2009) for details of the contingency graph construction.

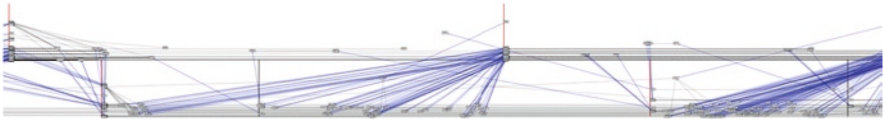

Fig. 16.3 A 20 min segment of an automatically generated contingency graph based on media dependencies

16.5.2.4 Contingency Graph Analysis

Selective Attention: Identifying Convergent Conclusions

The analysis begins with an important reference point in the interaction, a sequence of activity in which both participants express conclusions concerning the possible causes of ALS-PD. This episode takes place in a time span of approximately 10 min toward the end of the session. This portion of the record was selected as a starting point because it presented an opportunity to understand the conclusions expressed by participants as they began to summarize their ideas. The episode begins when P1 prompts for a conclusion, and ends when P1 and P2 explicitly agree that they are done. A content analysis of this segment of the interaction revealed an instance of convergence on "cycad usage" and an instance of divergence on "drinking water" as a causal agent for the disease. Using these terms we performed a search on the contingency graph to locate vertices containing matching references. Locating and highlighting vertices results in the identification of subgraphs of the larger contingency graph, which we call a trace.

Tracing: Subgraph Building Reveals Non Verbal Interaction Pattern

As noted above, one of the consistent concepts indicated in P1's argument during the concluding segment is that "drinking water" is one possible cause for the disease. In order to build an account of how this concept arose through the interaction, the contingency graph was queried to highlight acts that reference that text string and the contingencies between those acts. The graph revealed references to "drinking water" that were included in the information provided to P1 in relation to aluminum as a potential cause of the disease. A second query was formulated to capture acts that also referenced "aluminum", extending the trace. The resulting trace is summarized schematically in Fig. 16.4.

The graph revealed that in two particular instances P1 shares information with P2 related to the contamination of drinking water by aluminum. P2 performs a series of moves evidenced by clumps of move events in the contingency graph. These acts by P2 do not contain linguistic responses; only a series of moves (drag and drop acts) in the evidence map. This pattern is consistent throughout the remaining portions of the session. The trace shown in Fig. 16.4 could indicate that P2 is moving nodes around in order to see them, or to get them out of the way: dragging and dropping of

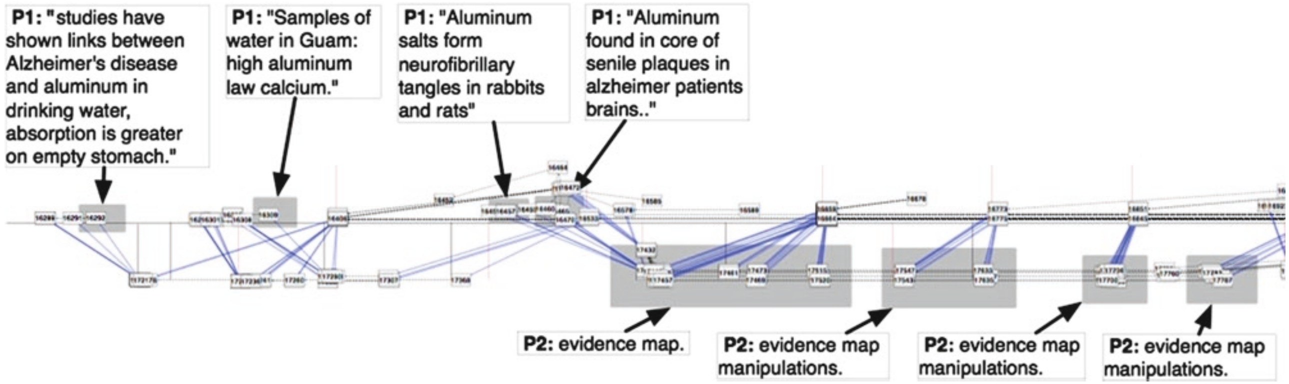

Fig. 16.4 Information sharing by P1 followed by systematic graph manipulations by P2 during the first hour of the session

graphical objects for these reasons is frequent. In this case however, the periodic-like pattern and density of P2's series of movements suggested more deliberate activity and induced us to explore the video record for these episodes to determine how these non-verbal manipulations might have influenced the interaction, especially with regard to the convergence identified above. The video shows that P2 is not randomly moving nodes around, but performing a series of evidence map *reconfigurations* to organize information previously shared during the session. After P1 contributes new information, P2 moves nodes to create spatially distinct groups that provide conceptual delineation. In addition to this spatial organization, both participants create links between nodes within groups that further clarify their inclusion in the group. (Their work will be illustrated in detail in the next section.)

Figure 16.5 illustrates the trace listed in Fig. 16.4 at a higher level of abstraction, as a series of uptake relations between episodic segments. Beginning at the left, P1 shares information containing a reference to aluminum in water as a contaminant in the first two segments [B1 & B3]. The third information-sharing event by P1 contains two references that correlate aluminum and neurological symptoms of ALS-PD [B6]. The reaction to the three sharing acts by P2 is shown as episodes of evidence map manipulations [B2, B4, B5 & B7–10]. Intersubjective uptake is indicated by P2's visual transformation of the shared information nodes and is followed by a series of intrasubjective transformative acts on the part of P2, who continually appropriates the relation-indicating power of the graphical nodes. The fact that there is very little related action on the part of P1 during these acts indicates that P2 is accountable for subsequent transformations. As shown on the far right of the diagram, intersubjective acts again occur as the concluding work segment (briefly discussed in section 0) is initiated [A1 & A2]. Closer examination of selected video segments using the contingency graph as an index reveals a purposed appropriation of the evidence map tool.

Micro Analysis: Indexing Video to Correlate Individual and Social Phenomena

The patterns represented in the contingency graph provided frames of reference and direct pointers, via timestamps, to relevant locations in the video record. More significantly, this framing made the interrelation between the two separate video streams (one stream for each participant) salient for determining the emergence of a shared representational practice.

As discussed above, there is a visible distinction in the participants' respective roles with regard to media coordinations: P2 does more graph related work and P1 does less action within the graph but expresses verbal articulations of hypotheses throughout the interaction. Their implicit role negotiation during joint problem solving developed early in the session and carried across different conceptual trajectories (Medina and Suthers 2008). For example, participants diverge on "aluminum", but converge on "cycad usage" as a cause of the disease. Next we describe key aspects of cycad convergence and how it is interactionally managed through representation.

Figure 16.6 shows the sequential introduction of cycad information and Figs. 16.7–16.9 show how this was done in the evidence map using screenshots.

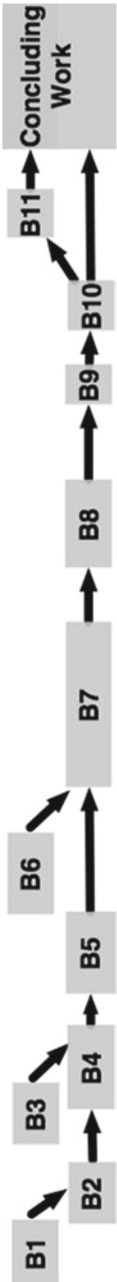

Fig. 16.5 High level view of uptake over the entire session

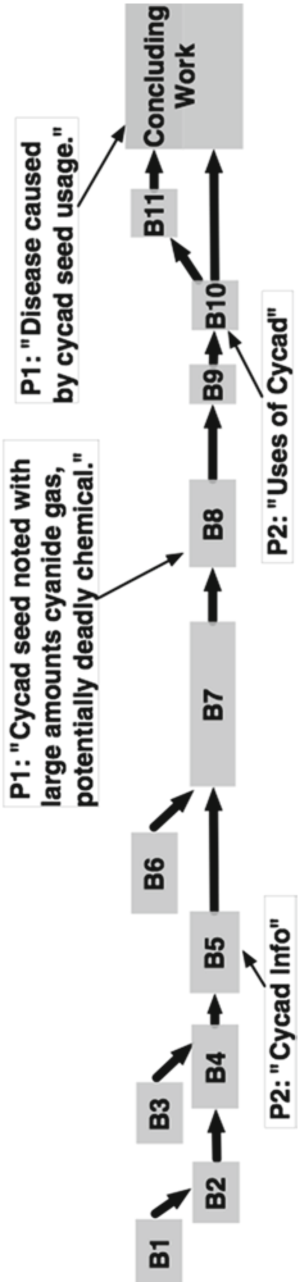

Fig. 16.6 Screen video is indexed in contingency graph (statements reference Figs. 16.7–16.10)

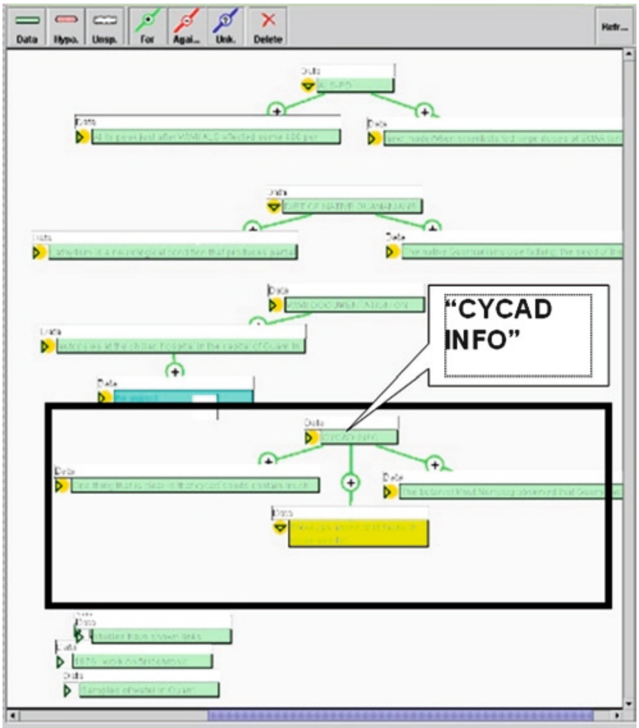

Fig. 16.7 P2 creates cycad representation

The reader is reminded that although participants are working in the same workspace, the environment updates changes in an asynchronous (delayed) manner, so the two participants' screens may diverge. We use bold rectangles and ellipses to help the reader track elements from one figure to the next

Following his own representational convention, P2 positions a label, CYCAD INFO, and three related data nodes into a configuration similar to other conceptually organized groups of nodes (identified by the rectangle in Fig. 16.7). Subsequently, P1 introduces a data node containing information about cycad seeds (identified by the ellipse, Fig. 16.8: time has elapsed, so P1's screen reflects the ongoing work of the two participants). In this context, a cycad related node is created and positioned in a somewhat arbitrary location with regard to the ongoing visual grouping. On receiving an update from P1 containing the cycad data (ellipse, Fig. 16.9), P2 reads the contents of the node, drags the node to a "member" position of the cycad conceptual grouping (rectangle, Fig. 16.9), and creates a "+" link between the node and the CYCAD INFO hub, further expressing its group membership (Fig. 16.9).

Subsequently, each participant brings "cycad usage" forward in distinct ways. P1 articulates cycad salience through a statement placed in a Hypothesis node, Disease caused by cycad seed usage (Fig. 16.10, left side), while P2 posts a short "themed" node expressing USES OF CYCAD (Fig. 16.10, *right side*). Each participant

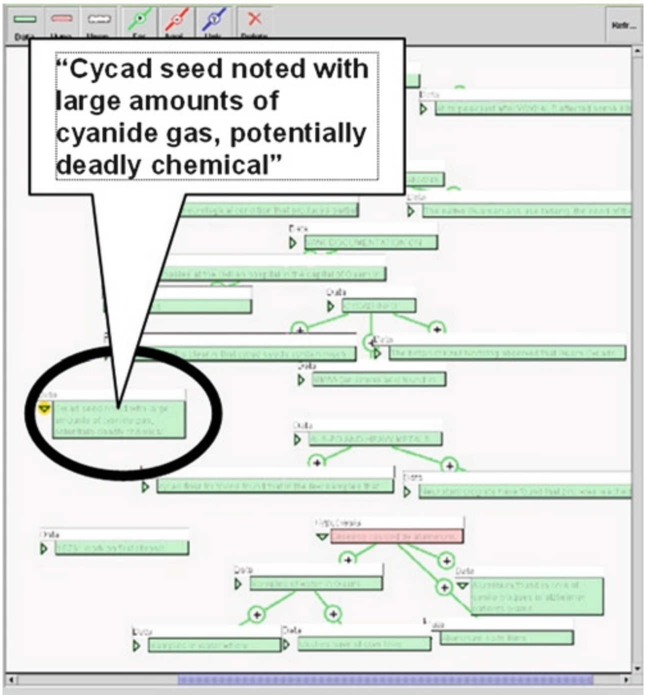

Fig. 16.8 P1 creates a cycad data node

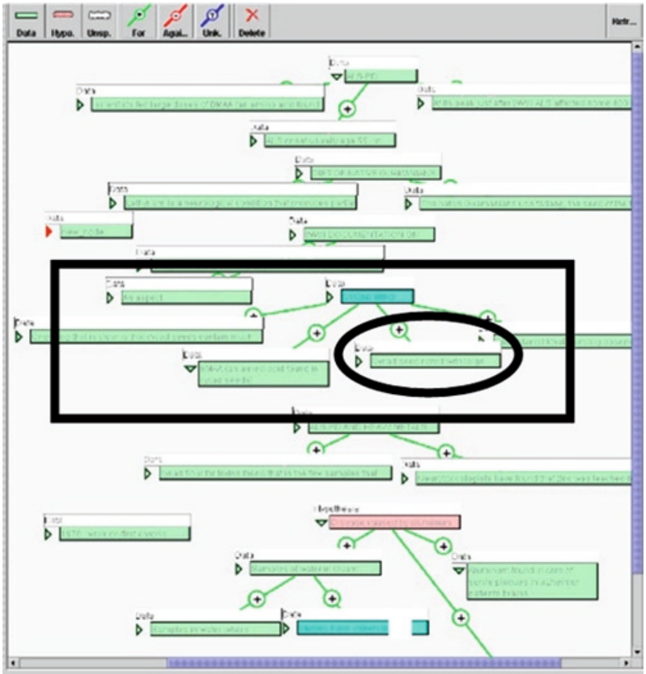

Fig. 16.9 P2 receives cycad data node from P1 (Fig. 16.8) and repositions and links into cycad group (Fig. 16.7)

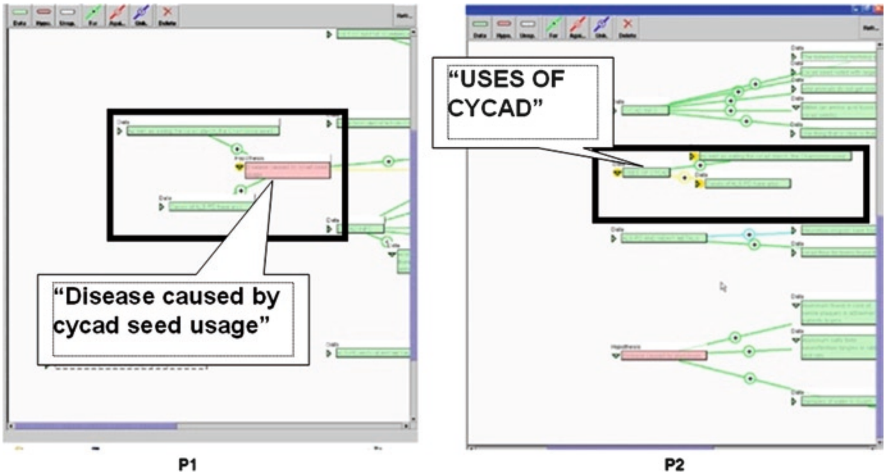

Fig. 16.10 P1 and P2 articulate new cycad groupings independently. (P2 has reorganized the graph into a horizontal layout format)

without knowledge of the other (they are "between" workspace updates) performs these respective acts in the shared workspace. They coincidentally indicate cycad usage at approximately the same time. In addition to posting her hypothesis node, P1 integrates it into the CYCAD INFO group configuration by creating four links to supporting data. P2 also groups and links data nodes to their expression (USES OF CYCAD). It is a mutual appropriation of a grouping practice. P1 and P2 both begin wrapping up their work within 5 min after this episode and thus initiate the concluding work episode presented above.

In summary, the initial indications that P2's evidence map manipulations were significant were substantiated through close examination of the interactional role of grouping practices. These practices were introduced and negotiated in alignment with the problem at hand. The inscriptional mechanisms carried the conceptual discourse in explicit and implicit ways.

16.5.2.5 Analytic Rationale for Using the Contingency Graph

The use of the contingency graph during this analysis supported flexible transitions between identification of macro interaction patterns and microanalysis of a series of graphical manipulations. Starting with analysis of short segments of interaction, it was possible to identify molecules of ideas. Using the contingency graph as an artifact encompassing the entire recorded interaction, it was then possible to identify individual and joint conceptual development. This, in turn, provided additional points that induced deeper investigation, transitioning back to analysis of micro segments. Understanding distributed interaction requires understanding both macro and micro phenomena in relation to each other. The contingency graph mediates between the two.

16.5.2.6 Scaling Up

Beyond the analytic applications we have already attempted, we envision using the contingency graph to scale up interaction analysis to larger data sets. For example, in order to understand learning in socio-technical networks (e.g., "online communities" and "virtual organizations"), we would like to explain how ideas develop as they move though socio-technical networks; how people acquire their roles in such networks; and how technological artifacts mediate these transformations of people and ideas as the artifacts move through networks and are themselves transformed. We are currently doing preliminary work in tracing out the pathways by which new social relationships are formed (Joseph et al. 2007; Suthers et al. 2009). It is our intention to generalize the contingency graph to an abstract transcript representation that supports a variety of analyses in larger scale social networks.

16.6 Remaining Challenges

In addition to generalizing and scaling up the approach, work remains for each of the challenges previously listed.

16.6.1 The Distributed Nature of the Data

The contingency graph gathers distributed data together into a single analytic artifact. In order to realize the benefits of such an artifact, we have translated from our log file representation to a contingency graph representation on an ad-hoc basis for the present analysis. More general tools are needed. The contingency graph should be defined as an abstract data type with an application program interface that can be used to write translators that import various log and sensor formats into the data type, and used to access the data type for analytic purposes. As a single representation of interaction, the contingency graph potentially enables analysts to "see" distributed interaction. However, the graph itself is an abstract structure. In order to realize this potential tools are needed to visualize and query that structure in useful ways (see Selective Attention, below).

16.6.2 The Contingent Nature of Human Behavior

The contingency graph helps address this challenge by making contingencies explicit, but at the same time the contingency graph confronts us with contingencies we may have previously ignored. Some colleagues have expressed apprehension about the potential complexity of contingency graphs, but the fact that complex and sometimes circumstantial evidence may be relevant to constructing valid accounts cannot be sidestepped. Instead, automated tools for identifying contingencies and querying contingency graphs (discussed later) are needed. Contingencies must include documentation of the evidence used to generate them, and analysts must be able to repudiate automatically identified contingencies as well as manually specify new ones.

We need to recognize that the record of proximal interaction may not capture all of the relevant contingencies, as some contingencies are based on events prior to the recorded interaction. An extension to our work could provide a means of introducing noninteractional data such as dispositions and prior knowledge into this analysis. Epistemological as well as methodological issues will need to be addressed, as this move involves combining two research traditions. In analogy to Latour's (2005) distinction between the "sociology of the social" and the "sociology of associations," dispositions and other prior individual differences belong to the "psychology of the psychological," while our approach might be called a "psychology of interactions" (see also Edwards 1997).

16.6.3 The Meaning of Nonverbal Behavior

The primary technical challenge in analyzing nonverbal behavior is to make the relevant media accessible. Ideally we would have an integrated environment in which one could access and view the media manipulations (graphical workspace manipulation, message postings, wiki edits, etc.) on demand (e.g., Brundell et al. 2008; Dyke and Lund 2009). The contingency graph can serve as an index and player synchronization device in such an environment, as well as being an analytic artifact in its own right.

Those technical accomplishments would make the relevant behavior available, but the hermeneutic problem would remain. By making the behavior available in a form that is amenable to inspection, collaborative interpretation including participants as well as other researchers becomes easier.

16.6.4 Selective Attention to Large Data Sets

Management of large data sets is in part a technical problem that will be addressed by the development of search, query and visualization tools aided by information filters. Other strategic solutions are specific to the purposes of the analysis. For example, in the analysis presented above we used visualization of the larger data set to leverage the human perceptual system's ability to detect interesting patterns that led to focused microanalysis of the original screen capture video. Other analyses may require tools for turning hypothesized interaction sequences into graph grammars to be searched for and counted (for example, Suthers et al. 2007).

In general, tools for managing large data sets should support movement between multiple levels of analysis, as discussed below.

16.6.5 Multi-scale Phenomena

Some of the most interesting challenges lie here. There are theoretical as well as empirical challenges. The nature of our data has led us to consider two questions of scale.

One question is the extent to which sequential analysis of interaction of the sort that is normally associated with microanalysis of face-to-face data (e.g., conversational transcripts or video) can be extended or scaled up along several dimensions: to interactions that are distributed across space (including locations and media), and across time (including asynchronous interaction); to include larger numbers of participants; and to longer time spans. This is partly an information processing and human-computer interaction problem of gathering the appropriate data and making it accessible to analysts in representations that make previously occult processes plainly visible. But there is also a problem concerning the adequacy of description: at what scales are interactive accounts productive, and at what point is it useful to shift to an aggregate level of description?

The other question concerns how implications for design can be drawn as we abstract away to aggregate phenomena. Even the largest aggregate phenomena fundamentally derive from individuals interacting with technological environments moment-to-moment, and making decisions at each moment that contribute to the aggregate results. These decisions are the point of contact between design of the technological environment and construction of the social reality. We are working on a theoretical account that bridges from these moments of experience and decision to other such moments at other times and places, aggregating to phenomena termed "social". Beginnings based on a cycle of "find/care/act-persist" are in (Joseph et al. 2007).

16.7 Broader Implications

Voluminous literatures exist for research that explores the relationship between preconditions or manipulated variables and learning outcomes, while black-boxing the processes by which this learning was accomplished. Such research is strong for hypothesis testing but weaker for discovery. As technological innovations proliferate and people's practices adapt to and adopt these innovations, we need to discover what is happening rather than to try to confirm what we have already guessed. This can be done only by making interaction visible: a challenge in complex and distributed environments. We have outlined an approach that "gathers together" interaction and makes possible tools that make it visible. We hope that the foregoing discussion has made clear how this approach enables a given analyst to understand learning in technology-mediated environments.

But there is a larger advantage to be realized as well. Progress in any scientific discipline requires that practitioners share conceptual vocabularies. Major advances in other scientific disciplines have been accompanied with representational advances, and shared instruments and representations mediate the daily work of scientific discourse (Latour 1990). Similarly, researchers studying learning in distributed and networked environments need shared ways of conceptualizing and representing what takes place in these environments to serve as the common foundation for our scientific and design discourse. Presently our community has neither a common representation of data nor a shared vocabulary to discuss it, so it is difficult to build on each other's work or to take advantage of the analysis tools built by different researchers. An abstract transcript format that captures relevant aspects of interaction in diverse media can serve as the basis for shared vocabulary when communicating with each other and for development and sharing of the software tools that are critically needed to scale up our analytic work. It is our intention that the abstract transcript describe here will be appropriated to support diverse analytic methods (ethnographic, sequential, statistical, etc.) consistent with the basic conceptual understanding of interaction outlined at the outset of this paper. To the extent that we succeed, the contingency graph can also serve as a boundary object (Star and Griesemer 1989) for discourse between disciplines that are addressing similar problems. A common abstract transcript can support and bridge between multiple theoretical perspectives and facilitate the application of different analytical methodologies and tools to complex data sets.

Acknowledgments The approach to analysis described in this paper was developed over several years in collaboration with Nathan Dwyer and Ravi Vatrapu. This work was supported by the National Science Foundation under CAREER award 0093505.

References

Barcellini, F., Détienne, F., Burkhardt, J.-M., & Sack, W. (2005). Thematic coherence and quotation practices in OSS design-oriented online discussions. In *Proceedings of the 2005 international ACM SIGGROUP conference on supporting group work (GROUP '05)* (pp. 177–186). Sanibel Island: ACM Press.

Berkowitz, M. W., & Gibbs, J. C. (1979). A preliminary manual for coding transactive features of dyadic discussion.

Blumer, H. (1969). *Symbolic interactionism*. Berkeley: University of California Press.

Bronckart, J. P. (1995). Theories of action, speech, natural language, and discourse. In J. V. Wertsch, P. D. Rio, & A. Alvarez (Eds.), *Sociocultural studies of mind* (pp. 75–91). New York: Cambridge University Press.

Brundell, P., Knight, D., Adolphs, S., Carter, R., Clarke, D., Crabtree, A., et al. (2008). The experience of using the digital replay system for social science research. In *Proceedings of the 4th International e-Social Science Conference*. University of Manchester, ESRC NCeSS.

Castells, M. (2001). *The internet galaxy: Reflections on the internet, business, and society*. New York: Oxford University Press.

Dillenbourg, P. (1999). What do you mean by "collaborative learning"? In P. Dillenbourg (Ed.), *Collaborative learning: Cognitive and computational approaches* (pp. 1–19). Amsterdam: Elsevier.

Dwyer, N., & Suthers, D. D. (2006). Consistent practices in artifact-mediated collaboration. *International Journal of Computer-Supported Collaborative Learning, 1*(4), 481–511.

Dyke, G., & Lund, K. (2009). Tatiana: An environment to support the CSCL analysis process. In C. O'Malley, D. D. Suthers, P. Reimann, & A. Dimitracopoulou (Eds.), *Computer supported collaborative learning practices: CSCL 2009 conference proceedings* (pp. 58–67). Rhodes: International Society of the Learning Sciences.

Edwards, D. (1997). *Discourse and cognition*. London: SAGE.

Garfinkel, H. (1967). *Studies in ethnomethodology*. Englewood Cliffs: Prentice-Hall.

Gibson, J. J. (1977). The theory of affordances. In R. Shaw & J. Bransford (Eds.), *Perceiving, acting, and knowing*. Hilsdale: Lawrence Erlbaum.

Hamilton, E., & Feenberg, A. (2005). The technical codes of online education. *E-Learning, 2*(2), 104–121.

Hmelo-Silver, C. E., Rebecca, J., Liu, L., & Chernobilsky, E. this volume. Representational tools for understanding complex computer-supported collaborative learning environments. In S. Puntambekar, G. Erkens, & C. E. Hmelo-Silver (Eds.), *Analyzing interactions in CSCL: Methods, approaches and issues* (pp. 83–105). New York: Springer.

Hutchins, E. (1995). *Cognition in the wild*. Cambridge: MIT Press.

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

Joseph, S., Lid, V., & Suthers, D. D. (2007). Transcendent communities. In C. Chinn, G. Erkens, & S. Puntambekar (Eds.), *The computer supported collaborative learning (CSCL) conference 2007* (pp. 317–319). New Brunswick: International Society of the Learning Sciences.

Koschmann, T. (2002). Dewey's contribution to the foundations of CSCL research. In *Proceedings of computer supported collaborative learning 2002* (pp. 17–22). Boulder: International Society of the Learning Sciences.

Koschmann, T., Stahl, G., & Zemel, A. (2007). The video analyst's manifesto (or the implications of Garfinkel's policies for studying practice within design-based research). In R. Goldman, R. Pea, B. Barron, & S. J. Derry (Eds.), *Video research in the learning sciences*. Mahwah: Lawrence Erlbaum.

Koschmann, T., Zemel, A., Conlee-Stevens, M., Young, N., Robbs, J., & Barnhart, A. (2005). How do people learn: Member's methods and communicative mediation. In R. Bromme, F. W. Hesse, & H. Spada (Eds.), *Barriers and biases in computer-mediated knowledge communication (and how they may be overcome)* (pp. 265–294). Amsterdam: Kluwer Academic.

Latour, B. (1990). Drawing things together. In M. Lynch & S. Woolgar (Eds.), *Representation in scientific practice*. Cambridge: MIT Press.

Latour, B. (2005). *Reassembing the social: An introduction to actor-network-theory*. New York: Oxford University Press.

Lee, A. S., & Baskerville, R. L. (2003). Generalizing generalizability in information systems research. *Information Systems Research, 14*(3), 221–243.

Marin, A., & Wellman, B. (2010). Social network analysis: An introduction. In P. Carrington & J. Scott (Eds.), *Handbook of social network analysis*. London: Sage.

Medina, R., Suthers, D., & Vatrapu, R. (2009). Inscriptions becoming representations. In C. O'Malley, P. Reimann, D. Suthers, & A. Dimitracopoulou (Eds.), *Computer supported collaborative learning practices: CSCL 2009 conference proceedings* (pp. 18–27). Rhodes: International Society of the Learning Sciences.

Medina, R., & Suthers, D. D. (2008). Bringing Representational Practice From Log to Light. In *International Conference for the Learning Sciences*. Utrecht.

Medina, R., & Suthers, D. D. (2009). Using a contingency graph to discover representational practices in an online collaborative environment. *Research and Practice in Technology Enhanced Learning, 4*(3), 281–305.

Norman, D. A. (1999). Affordance, conventions, and design. *Interactions, 6*, 38–42. May-June.

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

Reimann, P., Yacef, K., & Kay, J. this volume. Analyzing collaborative interactions with data mining methods for the benefit of learning. In S. Puntambekar, G. Erkens, & C. E. Hmelo-Silver (Eds.), *Analyzing interactions in CSCL: Methods, approaches and issues* (pp. 161–185). New York: Springer.

Renninger, K., & Shumar, W. (2002). *Building virtual communities*. Cambridge: Cambridge University Press.

Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language, 50*(4), 696–735.

Shipman, F. M., III, & McCall, R. (1994). *Supporting knowledge-base evolution with incremental formalization, CHI94* (pp. 285–291). Boston: ACM.

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

Star, S. L., & Griesemer, J. R. (1989). Institutional ecology, ‘translations’ and boundary objects: Amateurs and rofessionals in Berkeley’s museum of vertebrate zoology. *Social Studies of Science, 19*(3), 387–420.

Suthers, D. D. (2006a). A qualitative analysis of collaborative knowledge construction through shared representations. *Research and Practice in Technology Enhanced Learning, 1*(2), 1–28.

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

Suthers, D. D., Chu, K.-H., & Joseph, S. (2009). Bridging Socio-Technical Capital in an Online Learning Environment. In *Proceedings of the 42nd Hawai’i International Conference on the System Sciences* (HICSS-42), January 5–8, 2009, Waikoloa, Hawai’i (CD-ROM). New Brunswick: Institute of Electrical and Electronics Engineers, Inc. (IEEE).

Suthers, D. D., Dwyer, N., Medina, R., & Vatrapu, R. (2007a). A framework for eclectic analysis of collaborative interaction. In C. Chinn, G. Erkens, & S. Puntambekar (Eds.), *The computer supported collaborative learning (CSCL) conference 2007* (pp. 694–703). New Brunswick: International Society of the Learning Sciences.

Suthers, D. D., Dwyer, N., Vatrapu, R., & Medina, R. (2007b). An abstract transcript notation for analyzing interactional construction of meaning in online learning. In *Proceedings of the 40th Hawai’i International Conference on the System Sciences* (HICSS-34), (CD-ROM). New Brunswick: Institute of Electrical and Electronics Engineers, Inc. (IEEE).

Suthers, D. D., Harada, V. H., Doane, W. E. J., Yukawa, J., Harris, B., & Lid, V. (2004). Technology-supported systemic reform: An initial evaluation and reassessment. In *Paper presented at the proceedings of the sixth international conference of the learning sciences* (pp. 537–544). Santa Monica: International Society of the Learning Sciences.

Suthers, D. D., & Hundhausen, C. (2003). An experimental study of the effects of representational guidance on collaborative learning. *Journal of the Learning Sciences, 12*(2), 183–219.

Suthers, D. D., Hundhausen, C. D., & Girardeau, L. E. (2003). Comparing the roles of representations in face-to-face and online computer supported collaborative learning. *Computers and Education, 41*, 335–351.

Suthers, D. D., Medina, R., Vatrapu, R., & Dwyer, N. (2007c). Information sharing is incongruous with collaborative convergence: The case for interaction. In C. Chinn, G. Erkens, & S. Puntambekar (Eds.), *The computer supported collaborative learning (CSCL) conference 2007* (pp. 714–716). New Brunswick: International Society of the Learning Sciences.

Suthers, D. D., Vatrapu, R., Medina, R., Joseph, S., & Dwyer, N. (2008). Beyond threaded discussion: Representational guidance in asynchronous collaborative learning environments. *Computers and Education, 50*(4), 1103–1127.

Suthers, D. D., Yukawa, J., & Harada, V. H. (2007d). An activity system analysis of a tripartite technology-supported partnership for school reform. *Research and Practice in Technology Enhanced Learning, 2*(2), 1–29.

Teasley, S. D., & Roschelle, J. (1993). Constructing a joint problem space: The computer as a tool for sharing knowledge. In S. P. Lajoie & S. J. Derry (Eds.), *Computers as cognitive tools* (pp. 229–258). Hillsdale: Lawrence Erlbaum.

Chapter 17
Analyzing Collaborative Interactions Across Domains and Settings: An Adaptable Rating Scheme*

Nikol Rummel, Anne Deiglmayr, Hans Spada, George Kahrimanis, and Nikolaos Avouris

Abstract In this chapter we report on the development of a rating scheme for the analysis of collaborative process data, and on its implementation in diverse CSCL settings. The rating scheme is composed of nine dimensions measuring different aspects of collaboration quality: sustaining mutual understanding, dialogue management, information pooling, reaching consensus, task division, time management, technical coordination, reciprocal interaction, and individual task orientation. It can be applied to recordings (video, audio, screen recordings, or log data) of student interaction and does not necessarily require transcripts or written records. While the rating scheme was originally developed in the context of a specific CSCL setting (video-based interdisciplinary problem-solving in the medical domain; Meier et al. (2007)), we demonstrate in our chapter that it can successfully be adapted to other CSCL settings. First, we introduce the initial rating scheme and its dimensions. Next we describe the process of adapting it to data from a very different CSCL setting (chat-based interaction in computer science classes). We briefly report on a study that used the ratings of collaboration quality as basis for adaptive feedback to students on how to improve their collaboration. Finally, we describe how we have integrated our rating scheme with ActivityLens (Avouris et al. 2007), a software tool which allows for a combined analysis of multiple sources of data (e.g., logfiles, audio and video recordings). Several tool modifications were made to permit analysis of collaborative process data from yet another CSCL study in which high-school students collaborated face-to-face on solving algebra problems with support from an intelligent tutoring system. We conclude our chapter with a discussion of practical implications for practitioners who may wish to adapt and apply our rating scheme.

*Further information about the rating scheme and materials for rating can be obtained from the first author.

N. Rummel (✉)
Institute of Education, Ruhr-Universität Bochum, Universitätsstraße 150, 44801 Bochum, Germany
e-mail: nikol.rummel@rub.de

S. Puntambekar et al. (eds.), *Analyzing Interactions in CSCL*, Computer-Supported Collaborative Learning Series, DOI 10.1007/978-1-4419-7710-6_17,

In earlier work Meier et al. (2007) have developed a rating scheme that allows researchers to assess collaborative process quality in an economic fashion directly from video recordings. This method combines a qualitative with a quantitative approach, thereby avoiding some of the fallacies either one of them may have when used solely. The analysis approach is qualitative in that the rater does not count the occurrence of utterances of a particular type as is done in many coding systems (Strijbos et al. 2006). In contrast, the rater tries to understand and assess the interaction in its full complexity while watching the video recording. It is quantitative in that the assessment of collaboration quality is given in the form of quantitative ratings on several scales. In its original version (Meier et al. 2007), the rating scheme comprised the following nine dimensions: sustaining mutual understanding, dialogue management, information pooling, reaching consensus, task division, time management, technical coordination, reciprocal interaction, and individual task orientation. The rating scheme was initially developed, and successfully applied, in the context of a CSCL setting in which two partners from different domains (psychology and medicine) collaborated over a desktop videoconference system to solve (i.e. diagnose) complicated patient cases with combined psychological and physiological pathology.

The goal of two collaborative projects between psychologists at the University of Freiburg (Diziol, Deiglmayr (née Meier), Rummel, Spada) and computer scientists at the University of Patras, Greece (Avouris, Kahrimanis, Voyiatzaki), was to broaden the scope of the analysis method by adapting it to different CSCL settings and by integrating it with ActivityLens, a software tool developed by the Patras team which allows an integration of multiple sources of data (logfiles, audio and video recording, etc.; Avouris et al. 2007). The successful adaptation of the rating scheme not only provided us with a tool for analyzing new data sets, but also shows that the rating scheme's dimensions, and thus its underlying theoretical model, are capable of capturing the main aspects of collaboration quality across very different CSCL settings. In the first project we adapted the original rating scheme by Meier et al. (2007) to fit the collaborative process data (chat of dyads of computer science students, log data of shared whiteboard activities) of the Patras team (Voyiatzaki et al. 2008a). In addition, we explored the possibility of giving adaptive feedback to students based on an assessment of their collaboration (Meier et al. 2008). In the second project, we customized the ActivityLens software developed by the Patras group (Avouris et al. 2007) in order to analyze process data from another Freiburg project: In this project (Diziol et al. 2007; Rummel et al. 2010), students at the high school level collaborated in dyads supported by a computer-based tutoring system for mathematic instruction (Cognitive Tutor Algebra, © Carnegie Learning Inc.). Tutor log data and audio and screen capture of students' collaboration were combined using the ActivityLens tool, and an adapted version of the rating scheme was integrated in the software.

We will first summarize the original rating scheme. Then we will describe the adaptation of the rating scheme to the Patras data, the revised version of the rating scheme, and the development of a corresponding feedback scheme. Finally, we will shortly talk about the implementation of another version of the rating scheme in ActivityLens. The chapter will be concluded by a section in which we provide some practical advice for researchers who might want to use the rating scheme in analyzing their data.

17.1 The Original Rating Scheme[1]

Any researcher with interest in studying collaborative processes must decide which aspects of the collaborative process are relevant for the success of the collaboration and should therefore be observed. In principle, there are two complementary approaches to answering this question: The researcher can either start with data at hand or with a theoretical model in mind. In the development of the original rating scheme by Meier et al. (2007), a bottom–up and a top–down approach were combined in order to arrive at dimensions of good collaboration both grounded in the data and defined abstractly enough to be transferable to a broader range of computer-supported collaboration scenarios. More specifically, a qualitative content analysis of transcribed dialogue from an empirical study on computer supported collaborative problem solving was combined with theoretical considerations based on literature from areas such as collaborative learning, computer-mediated communication, and group decision making.

The development of the original rating scheme was embedded in a larger research project on instructional support for computer-supported, collaborative, interdisciplinary problem solving. The primary aim of this research project was to develop instructional measures in order to improve students' subsequent collaboration. Two studies were conducted within this project (Rummel and Spada 2005b; Rummel et al. 2009b). In both studies, dyads consisting of a medical student and a student of psychology collaborated via a desktop videoconferencing system. They worked on hypothetic patient cases that had been carefully designed to require the combined application of both medical and psychological expertise to be solved correctly. The desktop videoconferencing system allowed participants to see and hear each other while discussing the case. It included a shared workspace they could use to prepare a written joint solution as well as two individual text editors. In both studies dyads underwent a learning phase (experimental phase), during which they received instruction on solving a first patient case collaboratively. The effect of the instructional support was evaluated in a test phase. In this phase dyads in all conditions collaborated on the second patient case without additional support. The collaboration was videotaped. In addition, a post-test assessed individual knowledge about relevant aspects of collaboration in the present setting. The main goal of Study 1 was to evaluate two methods of instructional support that were implemented in the learning phase. In the model condition, participants observed a model collaboration in which two collaborators solved the first patient case. The model presentation consisted of recorded dialogue and animated text clips that allowed participants to follow the development of a model solution in the shared text editor. In the script condition, participants were provided with a script guiding them through their collaboration on the first case. Study 2 additionally investigated the effects of elaboration support provided in addition to model or script. Data from

[1] Some passages of the following text have been adapted from Meier et al. (2007). Copyright is held by the ISLS and permission was granted on March 15, 2009.

Study 1 were used in the development of the rating scheme's dimensions and data from Study 2 in its evaluation.

In the bottom–up approach, a multi-step analytical procedure building on the qualitative methodology developed by Mayring (2003) was followed in order to identify aspects of successful collaboration from the process data of Study 1 (Sosa y Fink 2003). Starting points were selected transcripts of students' collaboration in the test phase from Study 1. A qualitative content analysis of the transcripts led to a stepwise reduction of the material, through paraphrasing, elimination and generalization according to the rules established by Mayring (2003). Each step was documented and a final set of six categories was described and completed with anchoring examples (Sosa y Fink 2003). Subsequently, a complementary top–down approach was taken in order to refine these categories and arrive at process dimensions that would be relevant in a broader range of CSCL scenarios. Meier et al. (2007) reviewed literature on computer-supported collaborative learning and work, in order to identify aspects of successful collaboration under the conditions of video-mediated communication and complementary expertise. The search was guided by the results of the bottom–up approach.

17.1.1 Aspects of Collaboration Quality

The theoretical considerations that guided the refinement of the initial, empirically induced categories and the development of the final rating scheme addressed five broad aspects of the collaboration process: communication, joint information processing, coordination, interpersonal relationship, and individual motivation. In total, the final rating scheme comprised nine dimensions covering the essence of all empirically induced categories and all five aspects of collaboration considered important from a theoretical point of view. A more detailed description of the rating scheme can be found in Meier et al. (2007).

17.1.1.1 Communication

The success of any kind of collaborative activity depends, first of all, on effective communication. A "common ground" of mutually shared concepts, assumptions and expectations must be actively established and sustained during conversation (Clark 1996). To do so, speaker and listener must collaborate in ensuring understanding and in "grounding" their conversation (Clark and Brennan 1991). In particular, speakers must tailor their utterances to their partner's presumed knowledge level, a task that seems to be particularly hard to accomplish for experts talking to lay-persons or experts from other domains; they generally find it hard to ignore their own, specialized knowledge (Jucks et al. 2003; Nickerson 1999). The listener, on the other hand, is responsible for giving positive evidence of his or her understanding (Clark and Brennan 1991). In face-to-face conversation, this is usually achieved via eye contact or short verbal and nonverbal acknowledgments. However, in video-mediated and other computer-mediated communication, eye-contact usually is impossible and much non-verbal information is lost (Angiolillo et al. 1997; Rummel and Spada 2005a). Thus, participants need to employ more explicit feedback strategies, like verbal acknowledgements or paraphrases (Clark 1996), and have to check on their understanding more often than in face-to-face conversations (Anderson et al. 1997). As a prerequisite for a successful grounding process, participants need to ensure mutual attention (Clark 1996). A participant wishing to start a new episode of conversation has to check his or her partner's availability first. Further, turn-taking needs to be managed during conversation. Although turn-taking is governed by implicit rules (Sacks et al. 1974) that normally ensure relatively smooth transition in face-to-face communication, even small transmission delays in video-mediated communication can severely disrupt these implicit mechanisms. Thus, more explicit strategies have to be employed by participants, like handing over turns explicitly by asking a question or naming the next speaker (O'Conaill and Whittaker 1997). To summarize, communicators have to coordinate both the content and the process of their conversation.

Against this background, the first two dimensions of the rating scheme were defined as "sustaining mutual understanding" (which assessed grounding processes) and "dialogue management" (which assessed turn-taking and other aspects of coordinating the communication process).

17.1.1.2 Joint Information Processing

Collaborative problem solving requires participants to pool and process their complementary knowledge in a process of group-level information processing (Hinsz et al. 1997; Larson and Christensen 1993). Like face-to-face groups, partners in computer-supported collaboration must avoid falling prey to the general tendency of discussing primarily such pieces of information that were known to all group members from the start (Stasser and Titus 1985). This danger is even greater in interdisciplinary collaboration where the relevant information is distributed between experts (Rummel and Spada 2005a). Meta-knowledge about each others' knowledge bases and domains of expertise, that is, a transactive memory system (Wegner 1987), will facilitate the pooling of information (Larson and Christensen 1993; Moreland and Myaskovsky 2000; Stasser et al. 1995). In this way, participants are able to benefit from one another as a resource for problem solving and learning. Information can be pooled by eliciting information from one's partner or by externalizing one's own knowledge (Fischer and Mandl 2003). However, explanations must be given timely and at an appropriate level of elaboration in order to be helpful (Webb 1989). On the basis of the pooled information, collaborators must then reach a decision concerning the solution alternatives. This decision should be preceded by a process of critically evaluating the given information, collecting arguments for and against the options at hand, and critically discussing different perspectives (Tindale et al. 2003). Pressure towards group conformity (e.g., Janis 1982) as well as the tendency to avoid conflict and agree on a precipitate, illusory consensus can be counteracted by group norms valuing critical thinking (Postmes et al. 2001) and monitoring strategies emphasizing the quality of the group's solution (Tindale et al. 2003).

For the rating scheme, two separate dimensions were defined: "information pooling" (eliciting information and giving appropriate explanations) and "reaching consensus" (discussing and critically evaluating information in order to make a joint decision).

17.1.1.3 Coordination

Particularly in complex, non-routine tasks, the coordination of joint efforts is a crucial factor for the success of collaboration (Malone and Crowston 1990, 1994; Wittenbaum et al. 1998). Coordination is necessary because of interdependencies that arise when subtasks build upon each other, when time is limited, or when group members depend on the same resources (Malone and Crowston 1990, 1994). Discussing plans for how to approach a task and negotiating the joint efforts have been shown to be important for the quality of students' collaborative activities and outcomes (Barron 2000; Erkens et al. 2005). In planning their work, collaborators must take into account the nature of the task (Steiner 1972) as well as their individual resources and fields of expertise (Hermann et al. 2001). For divisible aspects of the task, individual work phases should be scheduled so that collaborators can bring their individual domain knowledge to bear, while joint phases are necessary for working on more integrative aspects of the task and ensuring a coherent joint solution (Hermann et al. 2001). In order to manage time constraints, a time schedule should be set up (Malone and Crowston 1994). In computer-mediated collaboration the aspect of technical coordination needs to be addressed in addition to task division and time management (Fischer and Mandl 2003). Shared applications, for example, constitute resource interdependencies that can be managed by setting up allocation rules (Malone and Crowston 1990).

Three dimensions represented the aspect of coordination in the rating scheme. The dimension of "task division" was defined to assess how well participants managed task–subtask dependencies. The dimension of "time management" assessed how participants coped with time constraints and the dimension of "technical coordination" assessed how they coped with technical interdependencies.

17.1.1.4 Interpersonal Relationship

Successful collaborative interactions are characterized by constructive interpersonal relationships. Collaborators often hold complementary knowledge that must be integrated in order to arrive at an optimal solution. They will be best able to do so in a relationship in which each of them has the same status, and in which perspectives are negotiable in a critical discussion (Dillenbourg 1999).

Dillenbourg has termed this a “symmetrical” relationship. Further, a respectful and polite tone of the conversation will help communicators to maintain face (i.e., feelings of self-worth and autonomy) and thus avoid negative emotions that would distract their attention from the task (Clark 1996). A constructive interpersonal relationship may be threatened by arising conflicts, e.g., if partners disagree on how to reach a shared goal. However, conflicts can promote productivity if managed constructively (Deutsch 2003). To achieve this, Deutsch advises collaborators to avoid stereotyped thinking and aggression, and instead to define conflicts as problems to be solved collaboratively.

In the rating scheme, one dimension was defined for this aspect of collaboration, reflecting Dillenbourg’s (1999) concept of the relational symmetry underlying collaborative interactions. This dimension, termed “reciprocal interaction,” denoted respectful, collaboratively oriented social interactions and the partners’ equality in contributing to problem solving and decision making, both of which should result from a symmetrical interpersonal relationship.

17.1.1.5 Motivation

Last but not least, the collaboration process will reflect participants’ individual motivation and their commitment to the collaborative task. Motivated participants will focus their attention on the task and co-orientate their actions around it, resulting in shared task alignment (Barron 2000). Possible motivation losses due to the group situation can be counteracted, for example, by strengthening individual accountability through mutual feedback (Johnson and Johnson 2003). Individual collaborators may employ volitional strategies to keep up a high level of expended effort in their contribution toward the joint task, including focusing their attention on solution-relevant information, keeping their environment free of distractions, or nurturing positive expectations regarding the collaborative outcome (Heckhausen 1989).

From observations of the dyads’ collaboration it became clear that participants sometimes differed substantially in their levels of task engagement, their willingness to spend effort on the task and to give feedback, and in their application of volitional strategies. Thus, the decision was made to assess participants’ motivation individually in the rating scheme. The resulting dimension of “individual task orientation” was rated separately for each participant.

17.1.2 Applying the Original Rating Scheme

The assessment of process quality requires a certain amount of interpretation by the rater, and thus might result in low objectivity if raters are not carefully trained. To counteract this problem, a rating handbook was written and used in rater training in order to standardize judgment and improve objectivity. The *rating* *handbook* contained a detailed description of each of the nine dimensions, along with illustrative examples and questions intended to guide raters' attention toward specific aspects of the collaborative process. The descriptions of the collaborative dimensions built on distinct behavioral acts that could be observed from video recordings of the collaboration process. Rating instructions were given by describing the "ideal" version of the dimension at hand, regarding both desirable characteristics that ought to be present as well as undesirable characteristics that ought to be absent. The raters' task was to judge to what extent the observed behavior matched the description in the rating handbook. In this way, the endpoints of the rating scales were defined as a "very good" match on the positive side and a "very bad" match on the negative side. Rating scales yield data that can be treated as approximately interval-level, in particular if "only the endpoints of the scale are named and denote the extremes of a continuum" (Wirtz and Caspar 2002, p. 124; translation by the authors). Therefore, only the endpoints of the rating scales were anchored verbally, while gradations were represented numerically.

The *rating sheet* listed ten scales, one for each of the first eight dimensions, and two scales for the dimension "individual task orientation", which was assessed separately for each member of the dyad. The scales had five steps that went from −2 (very bad) to +2 (very good). The rating sheet left some room under each dimension, and raters were encouraged to take notes on their impression of the dyad's performance in order to aid their memory and disambiguate the ratings.

17.1.3 Empirical Evaluation of the Rating Scheme

The rating scheme was evaluated with a new sample of dyads, the data of Study 2 (for more details, see Meier et al. 2007). Satisfactory inter-rater agreement could be achieved for most dimensions, and for the internal consistency of consecutive ratings for each dyad. The process ratings correlated moderately to highly, with the highest correlations between those dimensions designed to assess related concepts.

In addition, the rating scheme proved to be a sensitive measure for detecting effects of instructional support: The ratings revealed that the instructional methods employed in the learning phase of Study 2 had differential effects on the quality of the collaboration during the test phase (see Rummel et al. 2010, for a more detailed discussion). In another study (Rummel et al. 2007), the rating scheme was again applied with satisfactory inter-rater reliability, and, in that study, proved useful to assess the specific strengths and weaknesses of collaboration between persons at different levels of expertise in medicine and psychology.

Thus, within the setting studied by the Freiburg team (i.e. video-mediated collaborative problem-solving between psychologists and physicians) the instrument allowed for a sufficiently objective assessment of differential effects of different kinds of instruction, and of different levels of expertise on collaboration.

17.2 Adaptation of the Rating Scheme

The original rating scheme had been developed and tested in a specific CSCL setting. We believed, however, that the rating scheme's dimensions efficiently captured essentials of collaborative process quality in many areas of CSCL research and thus could be adapted for the analysis of data sets from other CSCL settings.

In a collaboration between the Freiburg team and the Patras team, we adapted the rating scheme to fit the Patras data. The data set consisted of computer logfiles from student dyads who jointly built the diagrammatic representation (flow-chart) of a classic algorithm (binary search) based on a written description of its properties and behaviors. The collaborative activity took place during typical laboratory classes and lasted approximately 45 min. Students collaborated through Syngero (Avouris et al. 2004), a network based synchronous collaborative drawing tool. Students could build their algorithm in a shared whiteboard, and at the same time communicate through a chat tool in the same window (see Fig. 17.1).

This setting differed significantly from the one the rating scheme had originally been developed for. Table 17.1 summarizes the main differences between the data of the Freiburg study for which the original rating scheme had been developed, and

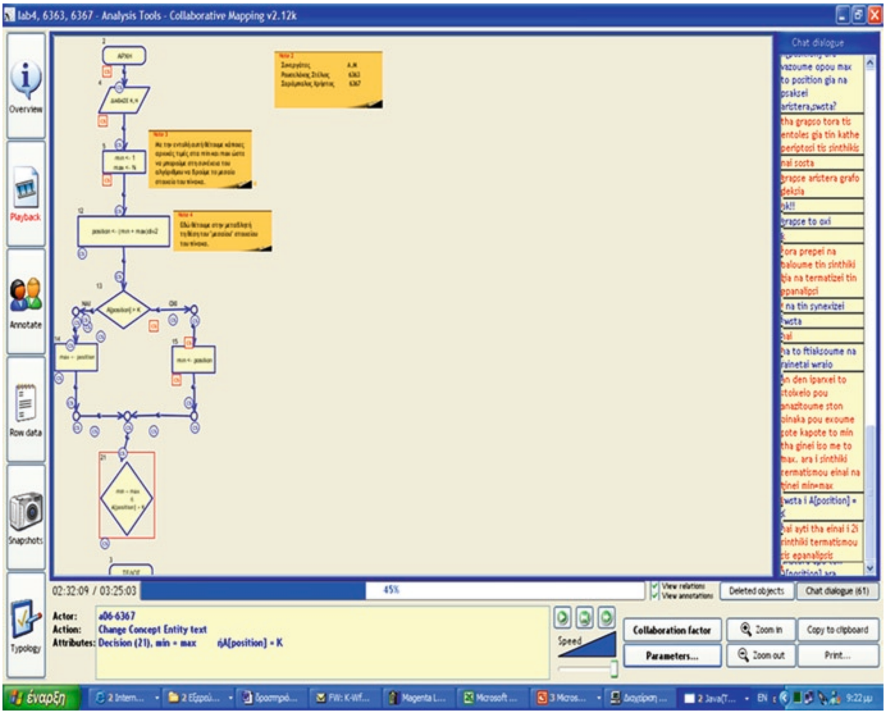

Fig. 17.1 Screenshot of the Synergo playback analysis tool, reproducing the contents of the shared whiteboard and chat produced by two students

Table 17.1 Main difference between data sets from Freiburg and Patras

	Freiburg	Patras
Domain	Medical decision making (diagnosing patients)	Computer programming (implementing algorithms)
CSCL setting and communication medium	Desktop-videoconferencing system with shared text editor	Synergo: shared whiteboard and chat tool in one window
Necessary knowledge resources	Knowledge-intensive task, collaborators with complementary knowledge	Task requires only basic knowledge; collaborators do not differ systematically in their prior knowledge

the data of the Patras studies for which it was adapted. Because the two setting were so different, we saw adapting the rating scheme to this novel CSCL setting, task, and sample as a major test of its theoretical validity and practical applicability.

17.2.1 Adaptation Process

The rating scheme was adapted to the specific task and setting described above by adjusting the selection and definition of the dimensions of the original scheme.

17.2.1.1 Redefining the Rating Dimensions

In the beginning of the adaptation process, we combined (top-down) considerations on how the differences in task, sample, and CSCL setting would change the meaning of successful collaboration in the Patras sample, with (bottom-up) best practice examples from the Patras data set. The sample consisted of ten student dyads who had participated in prior studies conducted by the Patras team (Voyiatzaki et al. 2008b). The definitions of all original rating dimensions were reformulated taking into account the situational constraints identified, and the observation of how successful dyads dealt with them in the case studies.

For example, an important aspect of one of the dimensions covering the communication aspect in the original rating scheme, “dialog management”, had been the coordination of turn-taking. In particular, transmission delays were common when students communicated over the videoconferencing system, requiring more explicit turn-taking than in normal face-to-face communication (O’Conaill and Whittaker 1997). However, the affordances of the chat tool in the Synergo setting were quite different. For example, the production costs in this medium are higher than in spoken communication, while, on the other hand, messages are reviewable, meaning that they can be inspected and referred to throughout the collaboration (Clark and Brennan 1991). In addition, actions in Synergo’s shared whiteboard could also be regarded as a form of communication (e.g. demonstrating a solution step instead of describing it verbally). In fact, inspection of the best practice examples showed that students switched naturally between these two channels of communication, for example by referring to actions in the whiteboard in their chat messages, by carrying out in the whiteboard suggestions that their partner had made in the chat, or by explaining verbally a part of the algorithm they had just constructed in the whiteboard. The chat messages alone, therefore, did not at first sight appear very coherent; however, taken together with the actions in the whiteboard, cross-references and coherence in dyads communication became apparent. As a consequence of both the theoretical considerations and the inspection of the data, we decided to build the adapted rating dimension around a concept we called "collaboration flow". This concept is more general, and thus has broader applicability, than the turn-taking concept. It refers to a coherent sequence of messages, both verbally and conveyed through actions, which build upon one another and thus enable the exchange and integration of knowledge and ideas in the collaborative problem solving process.

Further, students in the original scenario held complementary knowledge (the medical students had medical knowledge, the psychology student had clinical psychological knowledge) and therefore frequently exchanged their expertise-specific information during collaboration. This aspect had been part of the original "information pooling" dimension. Students in the Patras scenario, on the other hand, came from the same knowledge background (i.e. all of them were first year students of computer science). Accordingly, the case studies showed that plain information pooling was not very frequent in these dyads. Therefore, the focus of this dimension was shifted towards the explanations students provided for their actions and as a response to questions from their partner. The adapted dimension, accordingly, was named "knowledge exchange" rather than "information pooling".

All dimensions of the original rating scheme were changed in a process similar to the one just described. The resulting rating scheme included seven dimensions (Table 17.2): First of all, collaborators have to communicate successfully using chat as well as actions in the shared whiteboard. Partners have to maintain *collaboration flow*, that is, engage in a coherent exchange of information and maintain a joint focus. Further, they need to *sustain mutual understanding*, that is, work towards "common ground". Regarding work on the actual algorithm task, a very important dimension is that of *exchanging knowledge* (e.g. by giving self- and other-directed explanations). In addition, students have to engage in *argumentation* in order to ensure a good solution and to foster their own learning progress. This dimension refers to all activities involved in maintaining a critical discussion and double-checking the problem-solving process. To ensure a timely and orderly solution to the given problem, students also have to coordinate their collaboration well by *structuring the problem solving process*. Finally, students have to maintain a *cooperative orientation* (e.g., constructive handling of disagreements), and a high level of *task orientation* throughout their collaboration. Table 17.2 contrasts the original and the new dimensions of the rating scheme.

Table 17.2 Dimensions in the original and the adapted version of the rating scheme

Aspect of collaboration	Original rating scheme's dimensions	Adapted rating scheme's dimensions
Communication	Sustaining mutual understanding	Sustaining mutual understanding
	Dialog management	Collaboration flow
Joint information processing	Information pooling	Knowledge exchange
	Reaching consensus	Argumentation
Coordination	Task division	Structuring the problem solving process
	Time management	
	Technical coordination	
Interpersonal relationship	Reciprocal interaction	Cooperative orientation
Motivation	Individual task orientation	Individual task orientation

17.2.1.2 Adapting the Rating Handbook

The adapted rating handbook stated the scope and the operational definition of the adapted rating scheme's dimensions. As an example, Table 17.3 contrasts the operational definitions of the original *information pooling* dimension with the new *knowledge exchange* dimension. Both dimensions assess the aspect of joint information processing. However, in correspondence with the constraints of the task, the kinds of information processing that are in the focus of the two operational definitions, are different: While the main challenge students in the original scenario faced was to exchange their complementary knowledge resources, students in the new setting mainly had to understand the algorithm they were working with by formulating self- and other-directed explanations and by hypothesis generation and testing.

In addition, illustrative examples for each dimension were selected from the case studies. The *knowledge exchange* dimension, for example, was illustrated with several examples, including the positive example of successful knowledge exchange presented in Table 17.4.

As in the original scheme, each dimension was rated on a 5-point scale ranging from "very low" to "very high" collaboration quality on that dimension. The rating handbook was intended to be used in conjunction with a rater training that involved further illustration of the dimensions, and a more precise anchoring of the scales with the help of videos taken from the sample to be analyzed.

17.2.2 Empirical Evaluation of the Adapted Rating Scheme

An evaluation of the adapted rating scheme was undertaken in the context of an empirical study conducted jointly by the Freiburg team and the Patras team (Meier et al. 2008). The study was conducted in a classroom setting in actual programming lab courses at the University of Patras. Twenty-two dyads of first-year computer-science students collaborated on a first task in one of their weekly class meetings, received feedback on their collaboration according to experimental condition, and collaborated

Table 17.3 Definition and illustration of the original and the adapted version of the "information pooling" dimensions

	Original dimension	Adapted dimension
Name	Information pooling	Knowledge exchange
Statement of scope/ purpose	Collaborators have to pool, in particular, the unshared information (e.g. from their different materials, and from their different knowledge backgrounds) that each of them brings with them. In addition, they have to pool facts stated in the case description, and thus make it possible to integrate them in their shared solution	Collaborators make use of their knowledge resources in the process of solving their joint task. They learn from each others' explanations, by self-explaining or by explicit tutoring
Operational definition	Partners try to gather as many solution-relevant pieces of information as possible. New information is introduced in an elaborated way, for example by relating it to facts that have already been established, or by pointing out its relevance for the solution. In this way, the provider of the information ensures that it actually enters the problem solving process. Participants elicit domain-specific knowledge from their partner, using his or her expertise as a resource. At the same time they make sure that the aspects that are important from the perspective of their own domain are taken into account and they take on the task of clarifying any information needs that relate to their domain of expertise	Students ask each other for explanations, and give elaborated explanations. They use each other, as well as external sources, as a resource for obtaining the information they need to solve their joint task. In proceeding with their task, students explain to their partner or to themselves why they are doing what they are doing or suggesting to do

Table 17.4 Positive example of knowledge exchange from the rating handbook

00:25:12 RED: we want a variable that when the object of search is found to take a value and the algorithm to end
00:25:31 BLUE: aha m
00:25:43 RED: in order to go to the loop it has to be false otherwise it is out of the question
00:26:06 BLUE: a ok I got it
00:26:17 RED: so when it is found true...indeed we do it arbitrary because it comfortable for us
00:27:01 BLUE: L is the number we give in order to find it from the table
00:27:20 RED: l and r are like your min and max
00:28:15 BLUE: ok

Table 17.5 Measures of inter-rater reliability in an empirical evaluation of the rating scheme

Dimension	Intra-Class Correlation (ICC)
Sustaining mutual understanding	.79
Collaboration flow	.76
Knowledge exchange	.81
Argumentation	.77
Structuring the problem solving process	.70
Cooperative orientation	.82
Individual task orientation (dyad mean)	.71

on a second task in a following class meeting. The quality of students' collaboration on both tasks was rated using the adapted rating scheme. Inter-rater reliability, obtained by having a second rater assess one third of the dataset, was too low for some dimensions, despite being satisfactory for an aggregated measure of collaboration quality. As a consequence, the dimensions' definitions were fine-tuned and illustrated with additional examples. In addition, guidelines were defined for some uncommon observations which did not straightforwardly relate to a dimension's definition, and added to the rating handbook. In a final test of the so refined rating scheme, a new rater who had not been involved in the development or adaption of the scheme was trained with the new rater handbook and illustrative data sets. Then, both an experienced rater and the newly trained rater assessed the collaboration quality of 34 dyads that were selected randomly from the participants of several Patras studies with a total sample size of 101 dyads. In this co-rated sample, measures of inter-rater reliability (intra-class correlation for absolute values) were satisfactory, i.e. ICC ≥.70 for all dimensions (Table 17.5).

17.2.3 Giving Adaptive Feedback Based on Collaboration Quality Assessment

While the main focus of this chapter is on methods for assessment of collaboration quality, the model of collaboration underlying the rating scheme could in fact also be used to instruct students on how to improve their collaboration. To enable instructors to give students adaptive feedback, we developed a feedback scheme to be used in combination with the rating scheme. The feedback scheme contains a generic description, a positive feedback module and a negative feedback module for each of the six dimensions. In this way, it can be used to give *generic feedback* by explaining important aspects of collaboration to students and thus informing them about general strategies for their own collaboration. However, it can also be used to give students *adaptive feedback* by informing them not only about important aspects of collaboration, but also about the particular strengths and weaknesses of their collaboration, and by providing practical advice for improvement on the weaker dimensions (generic descriptions *plus* positive feedback for particularly high-rated dimensions and negative feedback for particularly low-rated dimensions).

Table 17.6 Feedback modules for the dimension "knowledge exchange" as stated in the feedback scheme

Generic description	"It is important for you to learn from your partner's knowledge, and let him learn from you. Therefore, ask for explanations if you have not completely understood what your partner is doing, and be sure that you explain understandably your own actions and reasoning."
Positive feedback module	"Your activities show that you are putting effort into explaining to each other what you are doing. Keep up with this good practice!"
Negative feedback module	"Your activities show that you need to give more explanations of what you are doing in order to improve the quality of your collaboration and your joint solution."

Table 17.6 gives an example of the three feedback modules for the dimension of "knowledge exchange".

In the evaluation study described above (Meier et al. 2008), a human tutor assessed collaboration quality using the rating scheme, and then gave feedback assembled according to the feedback scheme. In this way, the feedback was based on the "profile" of high and low ratings achieved by a dyad and thus tailored to students' specific strengths and weaknesses.

17.3 Implementing the Rating Scheme in ActivityLens

The goal of a second collaborative project was to customize the ActivityLens software developed by the Patras group (Avouris et al. 2007) for the analysis of process data from another Freiburg project: In this project (Diziol et al. 2007; Rummel et al. 2010), pairs of high-school level students collaborated with a computer-based tutoring system for mathematics instruction (Cognitive Tutor Algebra, © Carnegie Learning Inc.). Both during instruction and in the test period, the software saved log data of students' actions; furthermore, all interactions were recorded with audio and screen capture. To analyze the collaborative process, all data sources were integrated in ActivityLens. This made it possible to apply another adapted version of the rating scheme to analyze students' interactions taking into account log data, audio recordings and screen capture.

17.3.1 Adaptation of ActivityLens

The ActivityLens tool (formerly known as Collaboration Analysis Tool, ColAT) developed by the Patras team (Avouris et al. 2007) allows the evaluation of students' interaction by integrating data from several sources. For instance, researchers can combine video and audio data with log data from students' collaboration. These data are synchronized, enabling one to "jump" to a particular sequence of the interaction based on log data. The interface is shown in Fig. 17.3. It consists of a media window to display the video, and a text window for the log files. The log file window comprises three spreadsheets to analyze interaction on three hierarchical levels. On the first level, single events of the interaction are visualized (this sheet contains the original logfiles). On the second level, the researcher can combine several events to a “task”. On the third level, these tasks can be further combined to “goals”. Thus, the spreadsheets allow the researcher to perform a hierarchical analysis, reaching from a fine-grained evaluation of single events on the first level to a coarse-grained evaluation of goals on the third level.

In collaboration with the Patras team, several tool modifications were made to permit evaluation of collaborative process data from the algebra project. As a first step, the log data produced by the Cognitive Tutor Algebra and maintained in a database had to be made compatible for use in ActivityLens. After the data were in a format suitable for ActivityLens, some modifications in the functionality of the tool had to be made in order to improve its usefulness for the rating activity. We will refer to the version of ActivityLens that resulted from these modifications as “ActivityLens Freiburg Interface” (AFI). AFI enables researchers to browse data in the database where the logs of the Cognitive Tutor are stored. All data are dispersed in various tables of an SQL Server database. AFI, using appropriate SQL queries, reads the database management system and presents the collected data in a legible way. In addition, AFI allows the filtering of sessions so that the researcher can focus on specific types of problems in a certain session. An instance of the AFI tool is shown in Fig. 17.2. Finally, the application of the spreadsheets in ActivityLens was adapted in order to integrate a version of the rating scheme. In the adapted tool, the first spreadsheet displayed the Cognitive Tutor log data, while the second and third spreadsheets were used for collaboration process analyses with the rating scheme (see Fig. 17.3). In this project the collaborative process was analyzed from two perspectives. The first perspective assessed the quality of the dyads’ problem-solving process during particularly challenging problem-solving steps. It evaluated whether students took advantage of the resources in the learning environment in order to solve the problems and to increase their learning. The second perspective concentrated on the interaction process throughout the problem and assessed the quality of students’ collaborative behavior in more general terms. For each perspective there were several rating dimensions. The adapted version of ActivityLens incorporated some extra functionalities especially designed in order to make this possible. For example, in the second and third spreadsheet, a new mechanism was developed so that researchers could make copies of events and apply different annotations to them. This functionality permitted to rate one sequence of collaborative problem-solving with regard to different dimensions.

17.3.2 Using ActivityLens to Analyze Process Data from Algebra Study

First, the AFI was used to create the log file corresponding to a video in a format that is readable for ActivityLens. Together with the video data, these logfiles were uploaded in ActivityLens and synchronized. Before the rating, the interaction was

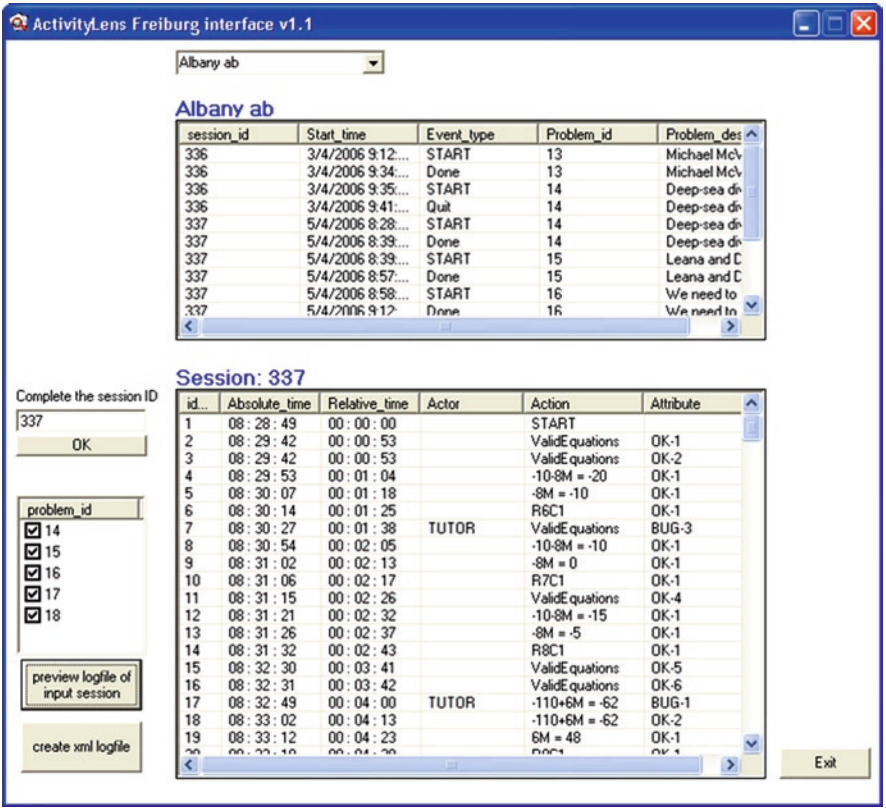

Fig. 17.2 Screenshot of the ActivityLens Freiburg Interface. Tool description: To extract a logfile from the database, the researchers chooses a student login (here Albany ab) from the *upper* drop-down menu. In the menu on the *left*, he can choose the mathematical problems solved by the student that should be included in the log file. The logfile is displayed in the large spreadsheet in the middle

divided into several sequences based on the log data. Figure 17.3 shows a screenshot of ActivitiyLens with data from the algebra project. The left window displayed the screen capture of students' collaboration. On the first spreadsheet, Cognitive Tutor logfiles were displayed. By marking an event, one could automatically jump to this sequence of the video file. On the second spreadsheet, the rater assessed students' interaction from the first perspective (problem-solving during difficult steps). Accordingly, assessment of students' interaction from the second perspective (general collaboration quality) was done on the third spreadsheet.

The application of the rating scheme with ActivityLens proved to be very efficient. In particular, it allowed taking into account different data sources (log data, audio, screen capture) and it made it easy to select specific sequences of the collaborative process.

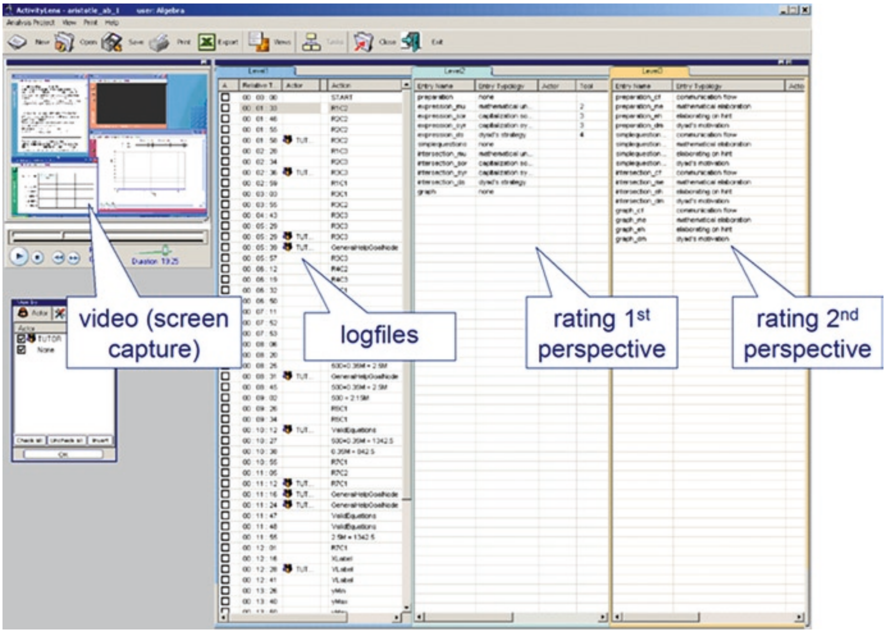

Fig. 17.3 Screenshot of the rating procedure with the ActivityLens tool

17.4 Implications for Practitioners

In the present chapter we have introduced a rating scheme comprising nine dimensions of collaboration quality in CSCL settings (Meier et al. 2007), and we have presented evidence for the objectivity and reliability of the initial rating scheme. We have described the adaptation process to the Patras CSCL setting, a setting very different from the one the original rating scheme had been developed for. For this adaptation we have also presented evaluation results supporting the objectivity of the rating scheme's application. We have demonstrated that the rating scheme can be applied efficiently to data from complex CSCL settings, whether it is used in a paper-based fashion or implemented in analysis software. As we have shown, the implementation of the rating scheme in the context of an analysis software like ActivityLens provides new possibilities: It enables researchers to integrate information from multiple data sources (e.g. log data, audio recordings, screen capture). On the basis of log data, particular parts of the collaboration process may be selected for deeper analysis. We have furthermore argued that process ratings could be used to provide adaptive feedback in order to improve learners' subsequent interactions.

In summary, we have shown that the dimensions of successful collaboration introduced in this chapter can be productively used to assess collaboration quality in a variety of CSCL settings, spanning diverse tasks, learning environments, subject domains, and student populations. When discussing the dimensions with both researchers and teachers, and when providing feedback to students based on the dimensions, we have found that they also seem to possess high face validity, which is an important asset in the practical application of an assessment instrument. Thus, we believe that rating schemes based on our dimensions may be successfully adapted to further CSCL settings in the future. However, the exact definition of the dimensions as well as the specific rating instructions will always have to be tailored to the specific task, setting and sample at hand. A rating scheme only allows judging the quality of the collaborative process against the relative standard of the best expectable collaboration within a given scenario. We have not attempted to define an "absolute" standard for the quality of collaboration in CSCL, nor do we believe that this is feasible. What distinguishes successful collaboration in a given setting will, of course, always depend on the task and the means available for solving it (e.g. the means of communication and collaboration afforded by the particular CSCL environment), but also on the goals of both the students and the teachers and/or experimenter. In the following section we summarize aspects that potential users of our rating scheme would need to consider when adapting it to their CSCL setting and applying it to their data set.

17.4.1 Adapting the Dimensions to Further Settings

A first step in adapting the rating scheme to your CSCL scenario is to redefine the meaning of the dimensions so they capture the essentials of successful collaboration in this setting. Ideally, you will already have some data at hand that was obtained in the same setting and format that you wish to analyze using the collaboration quality rating scheme. In our experience, it is then a fruitful exercise to extract some *"best practice"* and *"worst practice" examples* from the data, that is, identify instances of particularly successful and particularly unsuccessful collaboration. For example, in the adaptation of the original rating scheme to the Patras data set, we identified students who collaborated very well (or very badly) in general, and students who collaborated very well (or badly) concerning a specific aspect of collaboration (e.g. communication). The (bottom-up) analysis of particularly successful and particularly unsuccessful collaboration in the new setting should be combined with a (top-down) analysis of the specific affordances of the CSCL setting, for example, the collaborative task (e.g.: What kind of information processing is required? What subtasks need to be coordinated?), the CSCL environment (e.g.: What communication channels are available? How can students coordinate their work on the task?), or the composition of the group (e.g.: Do students have the same or similar knowledge background? What kind of interpersonal conflicts might arise during collaboration?). Based on these combined analyses, relevant dimensions can be selected and their scope and definition can be adapted.

A definition of each dimension should be stated explicitly in a *rating handbook*. Further, the rating handbook should provide raters with a standard of successful collaboration, for each dimension, against which the to-be-rated performance can be judged. The rating handbook serves to establish common ground between raters, and to train new raters. For both purposes, we also recommend to compile a collection of illustrative examples. For example, transcripts, excerpts or a library of video clips can illustrate relevant actions and interaction patterns, both desirable ones (e.g. critically discussing a proposal, or giving each other feedback) as well as undesirable ones (e.g. agreeing on a precipitate consensus, or neglecting time constraints). The use of examples from the videotapes (or chat and action logs) of the actual collaboration within the scenario to be studied has proven to be a particularly successful method of rater training. A video-based training makes it easy to demonstrate favorable behaviors and interaction patterns which are sometimes difficult to describe in a rating handbook. However, some dimensions of the rating scheme tap into long-term processes (e.g. the development of a consensus over a longer discussion; setting up a time and task schedule and monitoring adherence during the complete problem solving process), which cannot be illustrated by short video-clips or by transcript excerpts. For these dimensions, logfile charts representing work or discussion phases could be prepared. The video-clips and the charts could then also be used as the standards against which the videotaped interaction process is judged during the actual rating procedure. Raters also need an adequate amount of *background knowledge* about the specific demands of the task and the setting. Regarding the "sustaining mutual understanding" dimensions, for example, the rater should have enough background knowledge of the task domain in order to judge whether participants are able to reach common ground in their conversation, or whether they, for example, attach discrepant meanings to the same term. Thus, raters should use every opportunity to gather experience with the CSCL setting, for example by trying to solve the same task using the same CSCL tools as the students whose collaboration quality they are about to assess.

17.4.2 Concluding Thoughts on When to Apply a Rating Scheme in the Assessment of Collaboration Quality

Any researcher with interest in studying collaborative processes in CSCL has to answer two basic questions: (1) which dimensions of the collaborative process are relevant for assessing collaboration quality in a given CSCL setting? And (2) what kind of instrument or which methodology should be used for assessment? So far, this chapter has primarily addressed the first question. In concluding we would like to present some thoughts on the second, methodological question.

Many approaches toward assessing collaborative processes in CSCL follow a "coding and counting" rather than a rating approach. Coding schemes (e.g., Strijbos et al. 2006) are employed to assess the frequency of specific behavioral indicators or types of utterances, whereas a rating scheme focuses on a more direct assessment of process quality by judging the observed behaviors against a defined standard (Kerlinger and Lee 2000). Coding schemes have proven very useful in studies focusing on the relevance of specific indicators for the success of collaborative learning (for example, particular kinds of meta-cognitive statements, as studied by Kneser and Ploetzner [2001]). However, a general problem with these approaches is that the frequency of behavioral indicators often does not inform one about the success of collaboration (Rummel and Spada 2005a). For example, more coordinative utterances do not necessarily indicate better collaboration, because too much coordinative dialogue reduces the time available for the task itself. Too many coordinative utterances might even be an indicator of failed attempts to coordinate collaboration efficiently, and thus indicate ineffectual coordination. On the other hand, when applying a rating scheme, details of the collaboration process are lost due to the aggregation processes involved in rating process quality. Thus, rating schemes can be used to evaluate the quality of collaboration processes on a relatively global level, whereas coding schemes may be necessary for more fine-grained analyses. Further, coding schemes require written records of interaction, e.g. chat protocols. If no such records are available, a rating scheme is an economical solution because it does not require the transcription of dialogue but allows the researcher to work with video or audio recordings of the collaboration process. Finally, rating collaboration quality relies on an integration of multiple observations by the rater (Kerlinger and Lee 2000) and thus is more subjective than the application of a fine-grained coding scheme. As we have shown above, this problem can be solved by providing detailed rating handbooks and rater training; however, training raters may require more time and effort than training coders. In summary, a rating scheme is a good method if the goal is to evaluate the quality of collaboration on a relatively global level, based on complex data from multiple sources (e.g. video and audio; chat and whiteboard; action logs and audio), and if extensive rater training is feasible.

References

Anderson, A. H., O'Malley, C., Doherty-Sneddon, G., Langton, S., Newlands, A., Mullin, J., et al. (1997). The impact of VMC on collaborative problem solving: An analysis of task performance, communicative process, and user satisfaction. In K. E. Finn, A. J. Sellen, & S. B. Wilbur (Eds.), *Video-mediated communication* (pp. 133–156). Mahwah: Lawrence Erlbaum Associates.

Angiolillo, J. S., Blanchard, H. E., Israelski, E. W., & Mané, A. (1997). Technology constraints of videomediated communication. In K. E. Finn, A. J. Sellen, & S. B. Wilbur (Eds.), *Video-mediated communication* (pp. 51–74). Mahwah: Lawrence Erlbaum Associates.

Avouris, N., Fiotakis, G., Kahrimanis, G., Margaritis, M., & Komis, V. (2007). Beyond logging of fingertip actions: Analysis of collaborative learning using multiple sources of data. *Journal of Interactive Learning Research JILR, 18*(2), 231–250.

Avouris, N., Margaritis, M., & Komis, V. (2004). Modelling interaction during small-group synchronous problem solving activities: The Synergo approach. In *2nd International Workshop on Designing Computational Models of Collaborative Learning Interaction*, ITS 2004, Brazil.

Barron, B. (2000). Achieving coordination in collaborative problem-solving groups. *Journal of the Learning Sciences, 9*, 403–436.

Clark, H. H. (1996). *Using language*. Cambridge: Cambridge University Press.

Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. In L. B. Resnick, J. M. Levine, & S. D. Teasley (Eds.), *Perspectives on socially shared cognition* (pp. 127–148). Washington: American Psychological Association.

Deutsch, M. (2003). Cooperation and conflict: A personal perspective on the history of the social psychological study of conflict resolution. In M. A. West, D. Tjosvold, & K. G. Smith (Eds.), *International handbook of organizational teamwork and cooperative working* (pp. 9–43). Chichester: Wiley.

Dillenbourg, P. (1999). Introduction: What do you mean by "collaborative learning"? In P. Dillenbourg (Ed.), *Collaborative learning: Cognitive and computational approaches* (pp. 1–19). Amsterdam: Pergamon.

Diziol, D., Rummel, N., Spada, H., & McLaren, B. (2007). Promoting learning in mathematics: Script support for collaborative problem solving with the Cognitive Tutor Algebra. In Chinn, C. A., Erkens, G., & Puntambekar, S. (Eds.), *Mice, minds, and society. Proceedings of the Computer Supported Collaborative Learning (CSCL) Conference 2007 I* (Vol. 8, pp. 39–41). International Society of the Learning Sciences, Inc.

Erkens, G., Jaspers, J., Prangsma, M., & Kanselaar, G. (2005). Coordination processes in computer supported collaborative writing. *Computers in Human Behavior, 21*, 463–486.

Fischer, F., & Mandl, H. (2003). Being there or being where? Videoconferencing and cooperative learning. In H. van Oostendorp (Ed.), *Cognition in a digital world* (pp. 205–223). Mahwah: Lawrence Erlbaum Associates.

Heckhausen, H. (1989). *Motivation und Handeln. [Motivation and behavior]*. Berlin, Heidelberg, New York: Springer.

Hermann, F., Rummel, N., & Spada, H. (2001). Solving the case together: The challenge of net-based interdisciplinary collaboration. In P. Dillenbourg, A. Eurelings, & K. Hakkarainen (Eds.), *Proceedings of the first European conference on Computer-Supported Collaborative Learning* (pp. 293–300). Maastricht: McLuhan Institute.

Hinsz, V. B., Tindale, R. S., & Vollrath, D. A. (1997). The emerging conceptualization of groups as information processors. *Psychological Bulletin, 121*(1), 43–64.

Janis, I. L. (1982). *Groupthink*. Boston: Houghton Mifflin.

Johnson, D. W., & Johnson, R. T. (2003). Training for cooperative group work. In M. A. West, D. Tjosvold, & K. G. Smith (Eds.), *International handbook of organizational teamwork and cooperative working* (pp. 167–183). Chichester: Wiley.

Jucks, R., Bromme, R., & Runde, A. (2003). Audience Design von Experten in der netzgestützten Kommunikation: Die Rolle von Heuristiken über das geteilte Vorwissen. [Audience design of experts in net-based communication: The role of heuristics about shared knowledge]. *Zeitschrift für Psychologie, 211*(2), 60–74.

Kerlinger, F. N., & Lee, H. B. (2000). *Foundations of behavioral research*. Fort Worth: Harcourt College.

Kneser, C., & Ploetzner, R. (2001). Collaboration on the basis of complementary domain knowledge: Observed dialogue structures and their relation to learning success. *Learning and Instruction, 11*(1), 53–83.

Larson, J. R., & Christensen, C. (1993). Groups as problem-solving units: Toward a new meaning of social cognition. *The British Journal of Social Psychology, 32*, 5–30.

Mayring, P. (2003). *Qualitative Inhaltsanalyse*. Grundlagen und Techniken [Qualitative content analysis. Foundations and techniques. Weinheim: Beltz.

Malone, T. W., & Crowston, K. (1990). What is coordination theory and how can it help design cooperative work systems? In *Proceedings of the conference on computer-supported cooperative work, Los Angeles* (pp. 357–370). New York: ACM.

Malone, T. W., & Crowston, K. (1994). The interdisciplinary study of coordination. *ACM Computing Surveys, 26*(1), 87–119.

Meier, A., Spada, H., & Rummel, N. (2007). A rating scheme for assessing the quality of computer-supported collaboration processes. *International Journal of Computer-Supported Collaborative Learning, 2*(1), 63–86.

Meier, A., Voyiatzaki, E., Kahrimanis, G., Rummel, N., Spada, H., & Avouris, N. (2008). Teaching students how to improve their collaboration: Assessing collaboration quality and providing adaptive feedback in a CSCL setting. Paper presented as part of the symposium by Rummel, N., &

Weinberger, A. New challenges in CSCL: Towards adaptive script support. *Proceedings of the 2008 International Conference of the Learning Sciences*, Utrecht, The Netherlands.

Moreland, R. L., & Myaskovsky, L. (2000). Exploring the performance benefits of group training: Transactive memory or improved communication? *Organizational Behavior and Human Decision Processes, 82*(1), 117–133.

Nickerson, R. S. (1999). How we know—and sometimes misjudge—what others know: Imputing one's own knowledge to others. *Psychological Bulletin, 125*(6), 737–759.

O'Conaill, B., & Whittaker, S. (1997). Characterizing, predicting, and measuring video-mediated communication: A conversational approach. In K. E. Finn, A. J. Sellen, & S. B. Wilbur (Eds.), *Videomediated communication* (pp. 107–132). Mahwah: Lawrence Erlbaum Associates.

Postmes, T., Spears, R., & Cihangir, S. (2001). Quality of decision making and group norms. *Journal of Personality and Social Psychology, 80*(6), 918–930.

Rummel, N., Diziol, D., & Spada, H. (2010). Scripted collaborative learning with the Cognitive Tutor Algebra: An experimental classroom study. Manuscript under revision.

Rummel, N., Hauser, S., & Spada, H. (2007). How does net-based interdisciplinary collaboration change with growing domain expertise? In Chinn, C. A., Erkens, G., & Puntambekar, S. (Eds.), *Mice, minds, and sociecty. Proceedings of the Computer Supported Collaborative Learning (CSCL) Conference 2007* (pp. 611–620). International Society of the Learning Sciences, Inc. ISSN 1819-0146.

Rummel, N., & Spada, H. (2005a). Instructional support for collaboration in desktop videoconferencing settings: How it can be achieved and assessed. In R. Bromme, F. W. Hesse, & H. Spada (Eds.), *Barriers and biases in computer-mediated knowledge communication—and how they may be overcome* (pp. 59–88). Berlin, Heidelberg, New York: Springer.

Rummel, N., & Spada, H. (2005b). Learning to collaborate: An instructional approach to promoting collaborative problem-solving in computer-mediated settings. *Journal of the Learning Sciences, 14*(2), 201–241.

Rummel, N., Spada, H., & Hauser, S. (2009). Learning to collaborate from being scripted or from observing a model. *International Journal of Computer-Supported Collaborative Learning, 4*(1), 69–92.

Sacks, H., Schegloff, E., & Jefferson, G. (1974). A simplest systematic for the organization of turn-taking in conversation. *Language, 50*, 696–753.

Sosa y Fink, S. (2003). Merkmale gelungener Kooperation. Eine qualitative Analyse netzgestützter Zusammenarbeit. [Characteristics of successful cooperation. A qualitative analysis of net-based collaboration.] Diploma thesis, Albert-Ludwigs-Universität, Freiburg.

Stasser, G., Stewart, D., & Wittenbaum, G. (1995). Expert roles and information exchange during discussion: The importance of knowing who knows what. *Journal of Experimental Social Psychology, 31*, 244–265.

Stasser, G., & Titus, W. (1985). Pooling of unshared information in group decision making: Biased information sampling during group discussion. *Journal of Personality and Social Psychology, 48*, 1467–1478.

Steiner, I. D. (1972). *Group process and productivity*. New York: Academic.

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

Tindale, R. S., Kameda, T., & Hinsz, V. B. (2003). Group decision making. In M. A. Hogg & J. Cooper (Eds.), *Sage handbook of social psychology* (pp. 381–403). London: Sage.

Voyiatzaki, E., Meier, A., Kahrimanis, G., Rummel, N., Spada, H., & Avouris, N. (2008a). Rating the quality of collaboration during networked problem solving activities. In *Proceedings of the 6th International Conference on Networked Learning* (pp. 409–416).

Voyiatzaki, E., Papadakis, S., Rossiou, E., Avouris, N., Paparizzos, K., & Hadzilacos, T. (2008b). One size does not fit all: A case study of combining networked learning methods and tools. In *Proceedings of the 6th International Conference on Networked Learning, 5–6 May 2008*, Halkidiki, Greece. http://www.networkedlearningconference.org.uk/abstracts/PDFs/Voyiatzaki_847-848.pdf. Retrieved 5 November 2008.

Webb, N. M. (1989). Peer interaction and learning in small groups. *Int J Educ Res, 13*, 21–39.
Wegner, D. M. (1987). Transactive memory: A contemporary analysis of the group mind. In B. Mullen & G. R. Goethals (Eds.), *Theories of group behavior* (pp. 185–208). Berlin, Heidelberg, New York: Springer.
Wirtz, M., & Caspar, F. (2002). *Beurteilerübereinstimmung und Beurteilerreliabilität. [Inter-rater agreement and inter-rater reliability]*. Göttingen: Verlag für Psychologie.
Wittenbaum, G. M., Vaughan, S. I., & Stasser, G. (1998). Coordination in task performing groups. In R. S. Tindale et al. (Eds.), *Theory and research on small groups* (pp. 177–204). New York: Plenum.

Chapter 18
Analytical Frameworks for Group Interactions in CSCL Systems

Kristine Lund

18.1 Introduction

Quite generally, all the authors in this section of the book seem to agree that some type of active participation in CSCL situations may constitute a main factor for learning, but they differ on what is considered important to take into account, what to look at specifically and which methods to use. This is an unsurprising acknowledgement as such differences are usually based on former education, biases inherited from mentors and current interests. In this commentary, I hope to get past the obvious by comparing these authors' differences along the aforementioned dimensions. Our goal is to evaluate the current state of affairs—as reflected by these chosen articles—and set the agenda for some specific remaining challenges.

18.2 Objectives of Analytical Frameworks for Group Interactions

As is well known, researchers in CSCL have many purposes in designing analytical frameworks and in carrying out their analyses of group interactions. In what follows, I summarize the (multiple) objectives authors gave for their work as well as some of the general objectives they cited from the CSCL literature.

18.2.1 Linking Process Quality and Knowledge Construction

The general goal of Stegmann and Fischer (this volume) is to test various theoretical assumptions between specific qualities of text-based knowledge building processes in CSCL and successful knowledge construction. In order to meet that goal, they

K. Lund (✉)
ICAR Research Laboratory, CNRS, University of Lyon, Lyon, France
e-mail: kristine.lund@univ-lyon2.fr

S. Puntambekar et al. (eds.), *Analyzing Interactions in CSCL*, Computer-Supported Collaborative Learning Series, DOI 10.1007/978-1-4419-7710-6_18,

propose a multidimensional approach for the qualitative coding of online discussions (MAQCOD) that defines rules for segmentation, coding and high level data analysis applied to the coded data. High-level data analysis is defined as creating new understandings from coded and aggregated data such as quantifying or visualizing mutual influence.

18.2.2 Mediating and Transforming Learning and Teaching with Technology

Scanlon (this volume) cites a variety of purposes as illustrated by the literature: investigate the benefits of CSCL, determine the mechanisms of collaborative learning, gain evidence to improve software design, provide educational guidelines and design effective tasks for collaborative learning. One specific research goal stemming from her and her colleagues' work is to make sense of how technology rich settings determine new communicative spaces in which collaborative learning can take place.

18.2.3 Supporting Instructors

The objective of Gweon, Jun, Lee, Finger and Penstein Rosé (this volume) is to support instructors in making early-stage assessments of student group work by addressing the real problems that they face in their practice. Their goal is to develop a computer-based assessment framework for automatic and unobtrusive monitoring of group interactions in face-to-face group meetings. They also propose that the automatic assessment of group processes from face-to-face interactions recorded as digitized speech can be considered as a general model. The goal for their analytical framework is that it be used to study groups in contexts other than project-based learning for engineering education and they thus present their work as a roadmap for other researchers.

18.2.4 Measuring the Quality of Collaboration

The goal of the original analytical framework (Meier et al. 2007) as described in Rummel, Meier, Spada, Kahrimanis and Avouris (this volume), is to measure the quality of collaboration in groups. In general, the authors seek to improve students' collaboration quality, as measured by their analytical framework, by manipulating certain parameters (e.g. instructional type). There were several additional goals of the chapter. First, the authors wanted to broaden the scope of the analysis method (rating scheme) by adapting it to a new CSCL setting, which resulted in redefining some of their dimensions. They also sought to integrate the rating scheme into the existing software ActivityLens (Avouris et al. 2007), causing parts of the interface and functions to be reprogrammed in a way that modified the original concept of the software. A future goal is to use the collaboration assessments with the rating scheme to instruct students on how to improve their collaboration.

18.2.5 Defining the Process of Interaction Analysis

Martínez Monés, Dimitriadis and Harrer (this volume) propose a general interaction analysis model, which is used to classify and analyze the existing problems that arise when researchers try to include collaboration-analysis software tools in CSCL settings. Their ultimate goal is to provide solutions to these problems in order to accomplish better integration of such tools into mainstream CSCL practices. The target users in this case are not only researchers, but also teachers and students. These tools could be employed for different purposes, such as assessment, regulation or self-reflection.

18.2.6 Making Interaction Apparent

In general, the authors Suthers and Medina (this volume) have the objective of understanding how learning stems from the "entanglements" of the activities of multiple individuals in technology-mediated environments. They cite five analytical challenges that can be interpreted as goals they intend to meet. The main challenge, encompassing the others, is to make interaction apparent, given the multitude of contingencies between acts or events that are distributed across time, space and media. The specific challenges arise from the distributed nature of data, the complexity of contingencies (and uptakes within them) of events in different modalities, the possible meanings of non verbal manipulations within graphic environments, scaling up to larger data sets and time scales, and finally not only understanding phenomena at a given temporal or social scale, but understanding relationships between phenomena across scales.

18.2.7 Summary of Objectives for Analytical Frameworks

The main goals of analytical frameworks for CSCL can be expressed in the following series of steps. Researchers may attempt to perform all steps in their research program or just concentrate on a specific step. The first step is to define ways to perform summative assessment (evaluation of the quality of outcomes) but also formative assessment (evaluation of the quality of collaborative processes and outcomes-in-progress). Next, the results of these assessments can be used to plan ways to ameliorate both outcomes and progress towards them and this can be done in a variety of ways. One way is to intervene on a specific parameter (e.g. instruction, technology, task, etc.) and set up a methodology with the goal of confirming hypotheses about correlations or causal relations between the characteristics of the parameter and the outcomes and/or processes. Sometimes attempts are made to automate parts of the methodology (e.g. coding). Next, particularly in experimental paradigms, ways of improving either outcomes or goals can be automated (e.g. chosen indicators can be used as feedback, given by technological means to participants). Finally, the whole CSCL context can be analyzed (CSCL software + interaction analysis software, etc.) in order to facilitate all of the previous steps (Martínez Monés et al. 2010.).

18.3 A Selection of Theoretical Assumptions Made by Authors of Analytical Frameworks for Group Interactions

Authors of research articles, in general, make a variety of assumptions while going about their work that reflect their beliefs about the world and their confidence in previous research results. In this commentary chapter, I study the chapters of the authors in this book section and make explicit their theoretical assumptions about the nature of collaborative learning and what they consider as important to know in relation to this.

All of the authors in this section seem to agree that there is a relation between group processes and outcomes. They differ however, in their opinion on which aspects of group processes are pertinent for successful collaboration and on how to qualify success in collaboration. These authors are also distinguished from one another by how they measure learning or whether they even choose to do so, at times favoring the study of other aspects of the collaboration that may be conceptualized by the authors as favoring learning, but not as *being* learning. For example, authors can document the types of interactive phenomena particular affordances of technology make possible between participants (see also Lund et al. 2007). When learning *is* measured, it may be done so in individual terms or in terms of the group (more rarely) and may focus on evaluating the product or alternatively focus on evaluating the processes (both individual and group) that *lead* to the evaluated product. In what follows I state for each chapter the theoretical assumptions the authors make, what related phenomena the authors believe are important to consider, and how they define and measure learning, if they do.

18.3.1 Positive Relations between Process Features and Knowledge Construction

Stegmann and Fischer (this volume) illustrate with the literature that CSCL has moved on from simple "coding and counting" methods to attempting to quantify certain qualities of collaborative learning processes.

Along with other CSCL researchers, these authors assume positive relations between specific features of the collaborative process and successful collaborative knowledge construction as measured individually. The features of the collaborative process that are presumed desirable are defined theoretically and may relate to communication norms and principles that are used as benchmarks for defining quality. For example, they hypothesize that completeness of single arguments during discussions is positively related to depth of cognitive processing of the individual or that longer argumentation sequences (those that include for example counter-arguments) are positively related to improved domain knowledge among participating individuals. Although Stegmann and Fischer study collaborative situations, any learning is measured as a type of individual gain, using experimental approaches, even if such individual gain is linked positively to desirable features of the collaborative process.

18.3.2 Positive and Negative Relations between Process Features and Context

In contrast, Scanlon's approach (this volume) assumes both positive and negative relations between the overall context and the ways in which learners collaborate; data is preferably taken in naturalistic rather than experimental settings in order to gather information on naturally occurring contexts. She argues that different participants (e.g. learners, teachers) have different expectations and understandings of learning situations and behave accordingly. It is these differences that we must understand by obtaining—in addition to data on outcomes—a detailed picture of how individuals interact within a group situation and how these interactions develop over time. Documenting the impact of technology on collaborative learning situations is especially of interest. Scanlon strives to understand learning in terms of processes and outcomes, but does not give details in her chapter on the results of how individual gains in outcomes were evaluated. Rather, she argues that particular technological affordances (e.g. eye contact made between participants over a provided video link) are positively correlated to specific types of discourse during problem-solving (e.g. meta-level discourse about a task such as hypothesis generation).

18.3.3 Predictive Relations between the Flow of Language Communication and Group Difficulties

Gweon, et al. (this volume) assume *predictive* relations between the flow of language communication and group difficulties. In their opinion, these difficulties can be revealed through interaction processes that occur during group work. They also believe that the study of language communication flowing between group members will give us insight into helping teachers support group functions as well as insight into designing more effective group learning environments. These authors take a socio-cultural perspective on learning (Lave and Wenger 1991) and define this as not only considering the personal skills set and conceptual knowledge individual students can gain from designing together, but also the experience they have of participating in the group, their practice of collaborative knowledge building and the way their position in the professional engineering community can positively change with experience.

One could argue that these authors put into play a series of events that in the end could result in measuring both individual and group learning in terms of processes that are informed by indicators (cf. Martínez Monés, et al., this volume). Gweon, et al. first ask teachers what they want to know about their groups. Interestingly, the teachers they questioned mentioned assessment categories dealing with process (e.g. goal setting, knowledge building) more often than assessment categories dealing with learning goals (e.g. skill application) or product (e.g. individual contribution). This prompted the authors to focus on process and not on learning goals or product. They remark that focusing on collaborative process skills is beneficial for two reasons: such skills are cross-disciplinary and by supporting the process, a teacher can influence both learning and product.

Gweon, et al. then decide whether the indicators for assessing process are observable and whether they can be reliably tracked by a human annotator. Finally they decide whether such information can be automatically tracked using machine learning techniques in order to produce a summary report for teachers. The summary report is meant to help teachers choose where to intervene during group work and for what reason. The research is not yet at this stage, but ultimately, if learning is defined as based on the indicators of good group work that teachers are looking for (e.g. setting goals, making progress, building knowledge, all both as a group or individually), then the summary report itself can be a measure of competently engaging (or not) in these activities (if the report has been demonstrated to be reliable). These processes could then be correlated to individual or group learning outcomes, in this particular case: the quality of the collaboratively designed project.

18.3.4 Definitive Relations between Nine Different Dimensions of the Collaborative Process and the Quality of Collaboration

Rummel, et al. (this volume) build their analytical framework on a variety of assumptions about collaborative learning based on previous research results from their own work as well as the literature. These results deal with five broad aspects of the collaboration process: communication, joint information processing, coordination, interpersonal relationship and individual motivation. The authors claim that the dimensions of their rating scheme (rated from −2: very bad to +2: very good) and its underlying theoretical model make it possible to capture the main aspects of collaboration quality across very different CSCL settings. Each set of assumptions translates into a particular rating scheme dimension. For example, the dimension sustaining mutual understanding (the extent to which this occurs during collaboration for a particular dyad) is based in part on the assumption that any kind of collaborative activity depends on effective communication and on establishing and sustaining a common ground of shared concepts and expectations (Clark 1996).

Each of their nine original dimensions (sustaining mutual understanding, dialog management, information pooling, reaching consensus, task division, time management, technical coordination, reciprocal interaction and individual task orientation) are based in the same way on a set of theoretical considerations and research results arguing that each dimension is an integral part of the quality of collaboration and should be measured. Almost all measures (the actual rating of each dimension from −2 to +2) are carried out on the level of the dyad; one measure is done at the level of the individual (two scales for each individual task orientation).

The authors mention previous results that show how the ratings revealed two relationships. Firstly, instructional methods have differential effects on the quality of collaboration and secondly, specific strengths and weaknesses of collaboration can be assessed between people having different levels of expertise in medicine and psychology. One finds implicit or explicit links to learning within the statements of scope/purpose and operational definition of particular dimensions. For example, in the (adapted) dimension of knowledge exchange, the statement of scope claims that collaborators "learn from each other's explanations, by self-explaining or by explicit tutoring". Although collaboration itself is the domain of learning in these authors' studies, the reference to learning from each other's explanations in the dimension of knowledge exchange implies learning in a different specific domain of knowledge. In their studies, the task was very challenging. Dyads consisting of two types of experts elaborated joint solutions on diagnoses and therapy plans for complex patient cases. Only knowledge on collaboration was pre and post-tested, not knowledge about psychology and medicine. This latter knowledge could be implicitly evaluated through addressing the quality of the joint solutions, but this proved quite difficult and in their original study, no substantial correlations between process ratings and solution quality were found.

18.3.5 Interdependent Relations between the CSCL Environment, Its Context of Use and the Interaction Analysis Tools

The authors Martínez Monés et al. suggest what is needed in our field in order that interaction analysis tools are profitably used in CSCL settings. They propose that we build a shared conceptualization of the interaction analysis process because there are *interdependent* relations between the CSCL environment, its context of use and the interaction analysis tools employed to answer research questions. In other words, the study of collaborative learning needs to consider interaction analysis as a main input to any process that tries to reflect on or to intervene in collaborative learning, such as evaluation or regulation, etc.

They reviewed eight projects and identified three levels of problems that researchers encounter when they try to use IA tools in CSCL situations: application level, architecture level and design level. The specific problems at the application level are that (1) data logs are not provided or are insufficient, (2) data is not documented and (3) it's not possible to process the data. The problems at the architecture level are that (1) not all relevant data is captured, (2) data are not synchronized, (3) data formats are not compatible and (4) the architecture does not allow the gathering of crucial data. Finally the problems at the design level are that (1) the design did not take into account the need to analyze data, (2) the intended object of analysis has not been integrated into the teaching practice and (3) the IA tool constrains the choice for the CSCL application.

Solutions to these problems need to be found if IA tools are to be used in CSCL settings in order to measure learning gains. In regards to the nature of the problems described above, solutions can either be proposed *during* the design process of the learning activity or *in relation* to the learning tools and the analysis tools. For these authors, learning or other phenomena potentially related to learning is measured in terms of *indicators* that questioners (researchers, teachers or learners) choose to analyze in order to answer their questions. There are many types of indicators (cf. also Djouad et al. 2009; Choquet 2007), but an example of a calculable one is "who is sending messages to whom". This particular indicator can be used to perform social network analysis. In some cases, the indicator can be compared to a norm such as "less than 10% participation of one student is too low for good collaboration". Such information sets the stage for correlating individual participation rates to individual or group learning outcomes.

18.3.6 Contingency Relations between the Trajectories of Intra- and Inter-subjective Meaning Making and the Media Participants Use

The authors Suthers and Medina dedicate a whole section of their chapter to theoretical assumptions. They give a specific definition of learning to be understood broadly as a judgment that some worthwhile transformation of individuals has taken place (Suthers 2006). Whereas the other chapter authors do not give a specific definition in the form of ("learning is…"), a similar one is implicit in most of their work. Suthers and Medina's approach is to work backwards from the recognition of individual learning and follow the interactional accomplishment of meaning-making that led to that learning. It follows that they assume *contingeny* relations between events that are accomplished by particular media and that are influenced by other events, also accomplished by media. Trajectories of intra- and inter-subjective meaning making and the media participants employ are (re)constructed by the authors in order to understand technology mediated collaborative and networked learning. In fact, events and the media that are used to carry them out are co-constitutive, both horizontally (between events) and vertically (between events and media). This approach is influenced by ideas in ethnomethodology of how social order is constructed but also influenced by the goal of designing for learning. These two influences lead to two tensions. Firstly, the authors point out that while all episodes of interaction are potentially interesting to an ethnomethodologist as instances of social organization, research in the learning sciences attempts to choose episodes where learning will take place rather than where it does not. Secondly, while ethnomethodological accounts seek only to be pertinent to members' accounts of their actions, the accounts these authors construct must be relevant not only to participants in the episodes analyzed, but also to the practices of educators and researchers on the periphery of these episodes.

In the example in their paper, Suthers and Medina measure learning by producing a detailed account of how participants converged on a conclusion, based on verbal (threaded discussion) and nonverbal (evidence map manipulations) interaction, as inferred from a contingency graph. Interestingly, the outcome measure for these authors is not formulated as individual knowledge acquisition, per se (tested for in an individual manner as for some of the other authors in this section of the book), but rather that the participants succeeded in coming to agreement on an explanation for a complex phenomenon. The very existence of this outcome measure is perceived as being indicative of learning. According to the authors, selection criteria of outcome measures will be a function of researchers' judgment concerning learning. The focus of analysis, however is the interactional account leading up to the desired outcome. Note that this interactional account can contain individuals reacting to their own events as well as to others'.

18.3.7 Summary of Selected Theoretical Assumptions

The authors of the analytical frameworks reviewed here assume a variety of relations between different types of elements that can be found in the general CSCL setting. If CSCL researchers treat CSCL participants' stream of activity (talk, movement, etc.) in a selective way (Kendon 1992), then the question becomes what in CSCL participants' behavior do they treat as focal (phenomenon being contextualized) and what do they treat as background (Goodwin and Duranti 1992). Authors assumed the following relations as focal points for research:

- *positive* relations between specific features of the collaborative process and successful collaborative knowledge construction as measured individually;
- *both positive and negative* relations between the overall context and the ways in which learners collaborate;
- *predictive* relations between the flow of language communication and group difficulties;

– *definitive* relations between nine different dimensions of the collaborative process and the quality of collaboration;
– *otherwise interdependent* relations between the CSCL environment, its context of use and the interaction analysis tools employed to answer research questions;
– *contingency* relations between the trajectories of intra- and inter-subjective meaning making and the media participants use.

Each of these focal relations could be further contextualized by specifying what is considered as "background" to them (the relevant field of action within which those relations are embedded). By summarizing the type of relations that authors assumed between elements in the CSCL setting and by considering what we, as a community mean by *setting*, we begin to see what other types of relations could potentially exist. Specifying them is left as an exercise to the reader.

Authors mainly defined learning as some type of individual gain, but often correlated or could correlate such gains with particular characteristics of the collaboration *or* with particular characteristics associated with a specific use of technology:

– the completeness of arguments;
– long argumentation sequences (e.g. including counter-argumentation);
– eye contact over a video link associated with meta-level discourse (e.g. hypothesis formation);
– assessing group work on the level of the process (e.g. goal setting);
– dimensions of collaborative quality that may include activities that favor learning as an integral part of their definition (e.g. interactive explanation, self-explanation);
– the unfolding of intra-and inter-subjective meaning-making over time as associated with specific uses of technology (e.g. how two participants converge on a conclusion through use of a threaded discussion and evidence map construction and manipulation);

Finally, obtaining the indicators issued from the collaboration situation that are used for measuring learning or measuring phenomena that by hypothesis or by assumption favor learning should be planned for in advance. In other words, the CSCL software and surrounding context should be able provide the information that is needed to either calculate or infer the indicator that will be used by the interaction analysis software.

18.3.8 Unit of Analysis: Individual, Group or Both?

It is also possible to contrast the mainly individualist definition of learning of our chapter authors (despite the fact that collaborative processes are also the focus of analysis) with the view that there can be a social theory of collaborative knowing where the focus is on how the group constructs intersubjective knowledge that emerges and appears within group discourse (Stahl 2006). The difference seems to be one of theoretical positioning on what is the most fruitful unit of analysis in order to address learning and/or the quality of collaborative situations. Stahl suggests that instead of measuring the effects of CSCL tools and collaborative experiences on the learning outcomes of individuals, the group could be viewed as an emergent phenomenon with its own set of ideas and behaviors. This assumes that collaboration is an interactive phenomenon with meaning-making properties at the group unit of analysis.

Additional authors—other than those in this book section—have also attempted to theorize processes of individual learning, collaborative learning or both with varying foci on the cognitive unit of analysis. Baker (2002) shows how varying degrees of alignment, symmetry and agreement between members of a dyad define eight forms of cooperative problem-solving interactions (e.g. dyads who are aligned, symmetrical and in agreement are doing co-construction; dyads who are not aligned, (arguing past each other), symmetrical and disagree, are doing apparent co-argumentation, etc.). Cress and Kimmerle (2008) use the wiki as an example to propose four forms of learning: two individual and two collaborative. Firstly, the two forms of individual learning are internal assimilation (quantitative individual learning) and internal accommodation (qualitative individual learning). Secondly, the two forms of collaborative knowledge building (carried out with respect to the wiki) are external assimilation (quantitative knowledge building), and external accommodation (qualitative knowledge building). Pea (1993) also borrows the concept of appropriation and proposes that it, together with meaning negotiation, are central to showing how crucial aspects of learning are built up through conversation between learners. Lund and Bécu-Robinault (2010) give specific examples of how during conversation, teachers and learners reformulate aspects of their own and each others' physics knowledge while moving between talk, gestures, drawings and manipulations. These multimodal reformulations that happen during interaction are the way in which participants appropriate concepts and accomplish intra- and inter-subjective meaning-making.

This small selection of research showing the differences in focus of cognitive unit of analysis for learning when people work on a task together, coupled with the foci of the authors in this section of the book already illustrate the field's diversity. As previously stated, researchers mostly define learning as an individual gain, but very often focus their analysis on the collaborative processes that dyads or large groups engage in. At the same time, there are many attempts to document and theorize these processes at the dyadic or group level, in some cases producing descriptive and potentially *predictive* models (e.g. Baker 2002).

18.4 A Selection of Methodological Assumptions Made by Authors of Analytical Frameworks for Group Interactions

In addition to assumptions from a theoretical point of view about what phenomena should be considered in regards to learning and what consequences this has for the choice of measuring learning, authors also have assumptions—closely related to theory—about how research should be carried out from a methodological point of view.

18.4.1 Origins of Research

Even the origins of the research process do not escape assumptions. Before looking at the authors' views in this section, let us cite Greeno (1998) on this subject. He favors bringing together theory-oriented and instrumental functions of research activities within a situative perspective that is implemented in design experiments (Brown 1992; Collins 1992). For Greeno, theory-oriented research is primarily organized by research questions geared to developing coherent concepts and explanations in a particular domain. Instrumental research—in the field of education—is primarily organized by problems concerning the improvement of learning environments, materials and instructional methods.

Similarly, Stegmann and Fischer state that scientific investigations begin with a problem in educational practice or a psychological or educational theory. For Martínez Monés et al., the analysis of participants' interactions is usually driven by a hypothesis to be confirmed or rejected by certain observations. This analysis is driven by the target group. For example, *researchers* might need to obtain a detailed view of the process, while *students* might benefit from visualizations of their participation in a forum as compared to other students in the same classroom. Rummel et al. formulate this idea as researchers having to decide which aspects of the collaborative process are relevant for the success of the collaboration and should therefore be observed. They argue that there are two complementary approaches: either start with data at hand or with a theoretical model in mind. Stegmann and Fischer write of moving from a problem or theory to empirical research questions and then to the formation of operational hypotheses. In contrast, Scanlon's opinion is that experimental approaches such as some of those alluded to above, are not sufficient to reveal the interactions that emerge between students, teachers, tasks and tools. In the same vein, Suthers and Medina remind us of the great quantities of research that explore the relationship between pre-conditions or manipulated variables and learning outcomes, while black-boxing the processes by which the learning was accomplished. In their opinion, we should be attempting to *discover* how peoples' practices are adapting to technological innovations rather than trying to confirm what we have already guessed. Such discovery can only be accomplished by making interaction visible. Recent software applications such as Tatiana allow for such visualizations of *coded or rated* interactions that are synchronized with primary data (Dyke et al. 2009).

I will now look at some other more specific methodological assumptions of our authors—always related to their theoretical perspective—such as tensions between methodological constraints and views of the interaction, performing comparisons of situations "with" and "without" technology, and a discussion on top-down *vs.* bottom-up approaches.

18.4.2 Tensions between Methodological Constraints and Views on Interaction

In previous work with other colleagues, Stegmann and Fischer decided to take individual learners as the unit of analysis because their main point of interest is individual knowledge acquisition (Stegmann et al. 2007). Their assumption is that such acquisition is a consequence of individual cognitive processes. Interestingly, they do admit that learners in a group cannot be regarded as mutually independent. But as this interdependence violates the random sample prerequisite of the statistical procedures used in their experimental psychology paradigm, they needed to find a solution so they can use their intended methodology. So they randomly selected one learner from each group in order to represent all the learners of that group. The reader is asked to consider how these authors' perspective changes on group processes as a result of choosing an individual to represent the group. If individuals are not mutually independent, how can we take into account their interdependence in our analyses (cf. Cress 2008)?

18.4.3 Is Technology a Variable?

In accordance with her view that participants will interpret their tasks differently based on their view of the context/setting, Scanlon favors using methodological approaches based on activity theory (Leontjev 1981) as they shift the attention of the researcher away from specific interactions with the computer to the activity seen as a whole. However, in previous much older studies, to which Scanlon makes reference in her chapter (Scanlon et al. 1993), she and her colleagues use a very different methodology. Of course, researchers can change methodological approaches and indeed particular methodological approaches only allow for specific types of research questions to be asked, so if a researcher wants to ask different types of research questions, oftentimes he or she must change methodological approaches.

This particular methodological change is interesting because it illustrates a tendency in the CSCL community. In the 1993 article, Scanlon and her colleagues compared problem solutions produced by pairs with technology and distance communication to problem solutions produced by pairs working only with paper and pencil. One can still see studies of this type today, but researchers have criticized this approach (cf. Baker 2004). In his argument, technology can simply not be a variable. When you "add" technology to a learning situation, you cannot keep the other parameters constant (e.g. task, nature of communication, etc.) in order to measure effect on learning gains, for example. The consequence of "adding technology" is that the nature of the very actions performed by participants is transformed. Action cannot be separated from the mediating tool and thus comparisons of "with" and "without" technology are not helpful because the use of tools qualitatively transforms the nature of the action (Vygotsky 1978; Rabardel 2001).

A good example of how computerizing a task can fundamentally change its nature can be found in the Cabrigéometre software (Baulac and Laborde 1989) that was developed in Grenoble, France. Cabrigéometre allows users to construct geometrical figures, similar to what one could do on paper with a compass, but differing in a very fundamental way. Users can "stretch" the geometrical forms within the interface and thus see which properties of the figure are invariant. This of course is not possible on paper and thus the simulation radically changes the task by giving students new access to geometrical concepts. Geometry becomes a different activity for students. This is of course, the whole point of educational technology, to be able to do things that were previously not possible and to understand how new mechanisms at play may affect learning. Comparison with "old" situations where the new activity was not possible are not helpful for this. Solomon's (2005) systemic research showed that attaining any worthwhile effects by including computers in the classroom necessitates the redesign of the entire learning environment. He suggests that once redesign takes place, the relations of attitudes, abilities, activities, perceptions and social relations with respect to achievement change in important ways.

Scanlon also writes that using technologically mediated collaboration changes the nature of the activity in ways we cannot predict. Such a statement reinforces Suthers and Medina's idea that discovering how practices are adapting to technology can only be accomplished by making interaction visible.

18.4.4 Top-down vs. Bottom Up?

Engeström (1987) reminds us of two omnipresent dangers in relation to gathering and selecting data. The first is blindly or intuitively selecting data without articulated justification. The second is using data in order to illustrate what the researcher is determined to prove. In both cases, criticisms focus on whether the data is representative or comprehensive. Engeström also reminds us of the necessity to answer the following three questions when constructing categories in a theoretical investigation: (1) how to *select* the data; (2) how to *process* the data into categories; and (3) how to bring these categories into worthwhile *contact with practice.*

For Stegmann and Fischer, once data has been gathered and selected, defining data structure (e.g. dimensions, categories within dimensions and the way of segmenting data) should be based on both top-down and bottom-up approaches. Dimensions are derived from theory and research questions are defined in terms of them, but categories within dimensions are also developed by exploring and interpreting the chosen data. While both are used, a top-down approach will be more important if the effects and relations between the variables being studied have already been theoretically conceptualized in a clear manner. On the other hand, a bottom-up approach will dominate if these effects and relations are still being initially explored in empirical analyses. In Stegmann and Fischer's view, in fact, the definition of the data structure is always a mixture of top-down and bottom-up approaches. In addition, they mention that some processes in CSCL only become visible when contextual information is taken into account, implying an analysis of features from the data set. Interestingly, they state that contextual variables are often not conceptualized by the theory of learning and knowledge building, but could that be what Engeström means by "bringing categories into worthwhile *contact with practice?*"

Rummel et al. also use both top-down and bottom-up approaches, but have applied them procedurally in the work described in their chapter: the latter first, followed by the former. They used qualitative content analysis built on an approach by Mayring (2003) to identify and refine aspects of successful collaboration in their data, ending with six defined categories, each with illustrative examples. They then further refined these categories according to relevant previous research results in the literature. The approach Rummel et al. took in adapting their rating scheme to a new CSCL situation also used a combination of top-down and bottom-up methods. They consulted the literature on how differences in task, sample and setting would change the meaning of a "successful" collaboration and they also analyzed "best practices", both as reflected in the new data set to which they sought to adapt their rating scheme.

Martínez Monés et al.'s approach is mainly bottom-up or data-driven as they illustrate problems and suggest solutions in direct relation with the research projects they analyzed. However their goal is that their theoretical analysis of the situation will help the field of CSCL to reflect and eventually adopt an approach usable by researchers, educational practitioners and technology providers.

Gweon, et al. look to practice for the construction of their categories (assessments that instructors habitually make in order to diagnose group difficulties) while as mentioned previously, Suthers and Medina are intent on making the accounts that they construct of learners' meaning-making relevant to the practices of educators and researchers, as well as those participating in the episodes studied.

18.4.5 Summary of Selected Methodological Assumptions

According to Salomon (2005), four considerations play a role in ones choice of a research approach: the paradigmatic assumptions one adopts, the perceived nature of the phenomenon to be studied, the questions to be asked about the phenomenon and the methodology to be used.

The authors reviewed here made assumptions about the motivations underlying scientific investigations, about which scientific paradigms were suitable for what type of research questions and how some paradigms were better suited than others for CSCL or at least for specific research questions within CSCL. In addition, some methodological constraints seem to compromise some researchers' visions of collaboration. Finally, the ways in which top-down and bottom-up techniques were conceptualized and used for constructing categories was discussed.

18.5 Remaining Challenges

In the section on objectives of analytical frameworks for CSCL, I suggested that the different goals found in our authors' chapters could be situated along a timeline of steps composed of assessing quality of outcomes and processes, improving outcomes and processes leading to outcomes (in relation to and in *harmony* with teachers' practice), optionally automating either ways of assessing or ways of intervening with the goal of improving outcomes or processes and finally taking a critical "meta-view" of the former steps by thinking about the interdependencies of CSCL systems and interaction analysis software and the contexts in which they are mobilized. This last step is not so much a re-occurring integral step of typical CSCL research, rather it is an opportunity for the CSCL field to perform self-reflection on the problems and solutions of applying analytical frameworks in context. It is placed last in the timeline as such self-reflection is based on knowledge about all the former steps.

All of the remaining challenges mentioned by the authors can be situated somewhere along this timeline.

18.5.1 Automating Assessing and/or Intervening

Gweon et al. remark that a next challenge for their research would be to automatically diagnose group problems from the instructor's point of view. The elements that instructors use for diagnosis consist of both directly observable and inferable evidence; the specific challenge lies in finding ways of automatically detecting the latter.

Stegmann and Fischer, also look forward to automating coding, but warn of automatic discourse analysis that assigns possibly incorrect codes due to superficial features (e.g. very long statements are assigned as "critical" or locating the word "because" leads to coding for a "grounded claim"). McLaren et al. (2007) show an example of how automatic on-line analysis of chat interactions in real-time can give instructors an overview of the learning situation and show where they might intervene. Stegmann and Fischer dream of a fully automated system—a collaboration script that could be used for interventions related to students' argument construction. Such a script could prompt students when they are not grounding and warranting their claims and gradually fade out as students gain proficiency.

18.5.2 Assessing Quality of Processes

The remaining challenges for Suthers and Medina revolve around taking their goals further (see the section on objectives). Firstly, tools are needed to visualize and query the contingency graph. Secondly, automated tools for identifying contingencies should be constructed, although in order to capture all relevant contingencies, data previous to the interaction should also be captured. Automatically generated contingencies should include documentation on how they were generated.

Researchers must be able to automatically reject the contingencies they don't agree with as well as manually specify others. Perhaps Tatiana (Dyke et al. 2009) could be leveraged for these activities? In Suthers and Medina's opinion, widening data gathering to include prior individual differences (e.g. knowledge, dispositions) calls for combining two research traditions ("psychology of the psychological" with "psychology of interactions") so epistemological and methodological issues need to be addressed. If such prior individual data were to be entered in Tatiana, contingencies could also be constructed between elements of the individually-oriented data and elements of the interaction. Another of Suthers and Medina's challenges concern the adequacy of description of accounts of interaction. At what scales (distributed across space and time) are such accounts productive and when is it useful to shift to an aggregate level of description?

Rummel et al. do not end their chapter with specific challenges. Rather they define the steps needed so that other researchers can adapt their rating scheme to new CSCL settings. However, a related paper (Rummel et al. 2007), used their rating scheme to show that growing domain expertise can have negative effects on distance collaboration amongst people from different domains (psychology and medicine) and thus a new challenge from this perspective would be to find ways to tailor support for such collaborating experts.

18.5.3 Research Questions, Data, Methods: Interdependencies

Scanlon points out that it is difficult in informal learning contexts to predict learning goals in advance. The challenge is thus one of developing appropriate methods for measuring learning when learning goals are elusive. In the case when learning goals *are* defined (based on curriculum in a classroom study) and pre and post tests can be developed to track learning gains, she states that the challenge is one of appropriate attribution of outcomes to particular sequences of interaction.

Interestingly, these two challenges can be viewed from the perspective of two different research paradigms. The latter challenge for Scanlon is situated within an experimental psychology methodology where the difficulty is controlling variables in order to attribute causality (e.g. isolating a discrete element of a complex educational situation and varying it in a second situation, while leaving all else unchanged to see if it makes a difference in some outcome). The former challenge recognizes the difficulty in checking for outcomes that are not known beforehand. In both cases, these challenges can be understood as calling into question the very goal of defining specific outcomes in advance (Engeström 2008) and using the hypothetico-deductive approach. In other words, could it be the case that another methodological approach (i.e. systemic) be more adapted to meeting these particular challenges (Salomon 2005), one which recognizes that elements are interdependent and inseparable in the sense that a change in one changes everything else and thus requires the study of patterns and not of single variables? If it is the case in the systemic view that variables, actions, events and constructs mutually define each other (Salomon 2005), then establishing causality between single variables and previously defined learning outcomes becomes somewhat non-sensical. However, as Salomon points out, systemic and analytical approaches can co-exist. Indeed they are complementary.

In their chapter, Suthers and Medina used visualization of a large data set to detect what they defined as interesting patterns that led to focused microanalysis of the original screen capture video. But their future goal is to automatically search for potential causal relations between desirable pre-defined hypothesized interaction sequences and outcomes of interest. The reader is asked to think about how systemic and analytical approaches are represented in this case.

18.5.4 Interdependencies of CSCL Systems and Interaction Analysis Software

Martínez-Monés et al. point to several issues that require further efforts. Increasing the interoperability of our data is primordial, but the complexity and the variety of CSCL settings makes this difficult to achieve. Suthers and Medina remark that because our community has neither a common representation of data nor a shared vocabulary to discuss it, it is difficult to both build on each other's work and use the analysis tools proposed by different researchers.

Martínez-Monés et al. propose that either CSCL researchers can define a minimal set of data that tools need to inter-operate or each tool can translate into and out from an agreed-upon generic format. In addition, they suggest that current interaction analysis tools and services can be improved upon and new ones can be developed so that activities such as assessment or monitoring are explicitly modeled and specified, together with the rest of the CSCL situation before the teaching-learning activity is implemented.

18.6 Conclusions

In this commentary, I reviewed the six analytical frameworks in this section of the book across three dimensions: purpose of the analysis, theoretical assumptions and methodological assumptions. These dimensions provided a lens through which I was able to compare and contrast approaches, perspectives and contributions.

Whatever the analytical framework, I hope to have shown throughout this commentary that the validity and range of application of each methodological approach is limited by the various assumptions that underlie it, by the data gathered and then chosen for analysis by the researcher and finally by the research questions that make sense to ask, given the above constraints. Such a claim recognizes the multivocality of our field and the necessity for complementary approaches.

Acknowledgments I would like to acknowledge colleagues Nathan Dwyer, Nancy Law, Carolyn Rosé, Dan Suthers and Chris Teplovs who participated with me in the organization of a series of international workshops that were based on the following dimensions: theoretical assumptions underlying an analysis of interactions, purpose of analysis, unit of interaction for analysis, data representations, analytic manipulations. These workshops were or will be held at ICLS 2008, CSCL 2009, the Alpine Rendez-vous 2009, ICLS 2010 and the Alpine Rendez-Vous 2011. A symposium has been submitted to CSCL 2011. I would also like to thank those authors of this book section who reacted to this commentary.

References

Avouris, N., Fiotakis, G., Kahrimanis, G., Margaritis, M., & Komis, V. (2007). Beyond logging of fingertip actions: Analysis of collaborative learning using multiple sources of data. *Journal of Interactive Learning Research (JILR), 18*(2), 231–250.

Baker, M. J. (2004). Recherches sur l'élaboration de connaissances dans le dialogue. Mémoire d'Habilitation à Diriger des Recherches (Université Nancy 2). Downloaded from: http://tel.archives-ouvertes.fr/docs/00/11/03/14/PDF/hdr.pdf (consulted on March 1st, 2010).

Baker, M. (2002). Forms of cooperation in dyadic problem-solving. In P. Salembier & T. H. Benchekron (Eds.), *Cooperation and complexity in sociotechnical systems* (Vol. 16, pp. 587–620). Lavoisier.

Baulac, Y., & Laborde, J. M. (1989). Cabri-géomètre: un logiciel pour un nouvel apprentissage de la géométrie. *Actes du Congrès Éducation et Informatique, 2*, 467–472.

Brown, A. L. (1992). Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom set tings. *Journal of the Learning Sciences, 2*, 141–178.

Clark, H. H. (1996). *Using language*. Cambridge: Cambridge University Press.

Choquet, C. (2007). Ingénierie et réingénierie des EIAH: l'approche REDiM, Mémoire d'Habilitation à Diriger des Recherches, Université du Maine.

Collins, A. (1992). Toward a design science of education. In E. Scanlon & T. O'Shea (Eds.), *New directions in educational technology* (pp. 15–22). Berlin: Springer.

Cress, U. (2008). The need for considering multilevel analysis in CSCL research: An appeal for more advanced statistical methods. *International Journal of Computer-Supported Collaborative Learning, 3*, 69–84.

Cress, U., & Kimmerle, J. (2008). A systemic and cognitive view on collaborative knowledge building with wikis. *International Journal of Computer-Supported Collaborative Learning, 3*, 105–122.

Dyke, G., Lund, K., & Girardot, J. -J., (2009). Tatiana: An environment to support the CSCL analysis process. *Proceedings of the International Conference on Computer Supported Collaborative Learning (CSCL) 2009 on CD-ROM*, June 13–18, University of the Aegan: Rhodes, Greece.

Djouad, T., Mille, A., Reffay, C., & Benmohamed, M. (2009). Ingénierie des indicateurs d'activités à partir de traces modélisées pour un Environnement Informatique d'Apprentissage Humain. In *STICEF: Sciences et Technologies de l´Information et de la Communication pour l´Éducation et la Formation* (Vol. 16, p. 43). 2009.

Engeström, Y. (1987). *Learning by expanding: An activity-theoretical approach to developmental research*. Helsinki: Ori enta-Konsultit.

Engeström, Y. (2008). From design experiments to formative interventions. Keynote At *The International Conference On Learning Sciences*, Utrecht, June 25.

Goodwin, C., & Duranti, A. (1992). Rethinking context: An introduction. In A. Duranti & C. Goodwin (Eds.), *Rethinking context: Language as an interactive phenomenon* (pp. 1–42). Cambridge: Cambridge University Press.

Greeno, J. G. (1998). The situativity of knowing, learning, and research. *The American Psychologist, 53*(1), 5–26.
Gweon, G., Jun, S., Lee, J., Finger, S., & Penstein Rosé, C. (this volume). A framework for assessment of student project groups on-line and off-line. In S. Puntambekar, G. Erkens, & C. Hmelo-Silver (Eds.), *Analyzing collaborative interactions in CSCL: Methods, approaches and issues* (pp. 293–317). New York: Springer.
Kendon, A. (1992). The negotiation of context in face-to-face interaction. In A. Duranti & C. Goodwin (Eds.), *Rethinking context: Language as an interactive phenomenon* (pp. 323–334). Cambridge: Cambridge University Press.
Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge: Cambridge University Press.
Leontjev, A. N. (1981). The problem of activity in psychology. In J. V. Werstch (Ed.), *The concept of activity in soviet psychology* (pp. 37–71). Armonk: Sharp.
Lund, K. & Bécu-Robinault, K. (2010). Learning physics as coherently packaging multiple sets of signs. In Gomez, K., Lyons, L., & Radinsky, J. (Eds.) Learning in the Disciplines: Proceedings of the 9th International Conference of the Learning Sciences (ICLS 2010) - Volume 1, Full Papers, pp. 404-411. International Society of the Learning Sciences: Chicago IL.
Lund, K., Molinari, G., Séjourné, A., & Baker, M. J. (2007). How do argumentation diagrams compare when student pairs use them as a means for debate or as a tool for representing debate?". *International Journal of Computer-Supported Collaborative Learning., 2*, 273–295.
Martínez Monés, A., Dimitriadis, Y., & Harrer, A. (this volume). An interaction aware design process for the integration of interaction analysis in mainstream CSCL practices. In S. Puntambekar, G. Erkens, & C. Hmelo-Silver (Eds.), *Analyzing collaborative interactions in CSCL: Methods, approaches and issues* (pp. 269–291). New York: Springer.
Mayring, P. (2003). Qualitative Inhaltsanalyse. Grundlagen und Techniken [Qualitative content analysis. Foundations and techniques]. Weinheim: Beltz.
Meier, A., Spada, H., & Rummel, N. (2007). A rating scheme for assessing the quality of computer-supported collaboration processes. *International Journal of Computer-Supported Collaborative Learning, 2*(1), 63–86.
McLaren, B., Scheuer, O., De Laat, M., Hever, R., de Groot, R., & Rosé, C. P. (2007). Using machine learning techniques to analyze and support mediation of student e-discussions, *Proceedings of Artificial Intelligence in Education: Building Technology Rich Learning Contexts That Work*. Frontiers in artificial intelligence and applications; Vol. 158, pp. 331–333.
Pea, R. D. (1993). Learning scientific concepts through material and social activities: Conversational analysis meets conceptual change. *Educational Psychologist, 28*(3), 265–277.
Rabardel, P. (2001). Instrument mediated activity in situations. In A. Blandford, J. Vanderdonckt, & P. Gray (Eds.), *People and computers XV -interactions without frontiers* (pp. 17–30). London: Springer.
Rummel, N., Meier, A., Diziol, D., & Spada, H. (this volume). Analyzing collaborative interactions across domains and settings: An adaptable rating scheme. In S. Puntambekar, G. Erkens, & C. Hmelo-Silver (Eds.), *Analyzing collaborative interactions in CSCL: Methods, approaches and issues* (pp. 367–390). New York: Springer.
Rummel, N., Hauser, S., & Spada, H. (2007). How does net-based interdisciplinary collaboration change with growing domain expertise? In C. A. Chinn, G. Erkens, & S. Puntambekar (Eds.), Mice, minds and sociecty. *Proceedings of the Computer Supported Collaborative Learning (CSCL) Conference 2007* (pp. 611–620). International Society of the Learning Sciences, Inc.
Salomon, G. (2005). Transcending the qualitative-quantitative debate: The analytic and systemic approaches to educational research. *Educational Researcher, 20*(6), 10–18.
Scanlon, E. (this volume). Analyzing collaborative interactions in multiple science learning contexts. In S. Puntambekar, G. Erkens, & C. Hmelo-Silver (Eds.), *Analyzing collaborative interactions in CSCL: Methods, approaches and issues* (pp. 319–339). New York: Springer.

Scanlon, E., O'Shea, T., Smith, R., O'Malley, C., & Taylor, J. (1993). Running in the rain-can a shared simulation help to decide? *Physics Education, 28*, 107–113.

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

Stegmann, K., Wecker, C., Weinberger, A., & Fischer, F. (2007). Collaborative argumentation and cognitive processing – an empirical study in a computer-supported collaborative learning environment. In C. Chinn, G. Erkens, & S. Puntambekar (Eds.), Proceedings *CSCL 2007 "Minds, Mice, and Society"*, Vol. 8, Pt. 2, pp. 661–670. July 16–July 21, Rutgers University, New Brunswick, NJ.

Stegmann, K., & Fischer, F. (this volume). Quantifying qualities: Multi dimensional analyses of text based communication in CSCL. In S. Puntambekar, G. Erkens, & C. Hmelo-Silver (Eds.), *Analyzing collaborative interactions in CSCL: Methods, approaches and issues* (pp. 247–268). New York: Springer.

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

Suthers, D., & Medina, R. (this volume). Tracing interaction in distributed collaborative learning. In S. Puntambekar, G. Erkens, & C. Hmelo-Silver (Eds.), *Analyzing collaborative interactions in CSCL: Methods, approaches and issues* (pp. 341–366). New York: Springer.

Vygotsky, L. (1978). In M. Cole, V. John-Steiner, S. Scribner, & E. Souberman (Eds.), *Mind in society: The development of higher psychological processes*. Cambridge: Cambridge University Press.

Index

S. Puntambekar et al. (eds.), *Analyzing Interactions in CSCL,* Computer-Supported Collaborative Learning Series, DOI 10.1007/978-1-4419-7710-6,