Multi-faceted Evaluation for Complex, Distributed Activities

Dennis C. Neale, John M. Carroll

Center for Human-Computer Interaction, Department of Computer Science, Virginia Tech

Abstract: Computer-supported cooperative learning presents challenges for evaluation methodology: Learning events and learning outcomes are dispersed in time and space, making causal relationships difficult to identify. We are developing techniques to address these challenges including systematic sampling, collation of multiple evaluation methods and data, and the use of collaborative critical incidents. In this paper we overview and discuss this emerging methodology.

Keywords: multi-user evaluation, distributed evaluation, qualitative and quantitative approaches, CSCL

Introduction

Evaluating outcomes associated with computer-supported cooperative learning (CSCL) is difficult for a variety of reasons. One must consider user interface usability issues, coordinated multi-user computer issues, learning efficacy in general, cooperative aspects of group learning specifically, and the larger context of the classroom(s) in which the previous issues are situated. Geographically displaced learning communities coordinating activities through computer networking technologies with individuals and subgroups often working and learning different things at different times and places presents even greater challenges for researchers assessing outcomes on students, teachers, and educational organizations. Specifically, problems result because measurement is dispersed in time and across place, and the subsequent evaluation stages as a result are more complicated because these activities occur across individuals and groups.

We have developed a general purpose multi-faced evaluation framework to address complex, distributed activities as they relate to multi-user (groupware) computer interfaces. We describe the framework in this paper specifically as it relates to the CSCL context in which we developed it. We outline the requirements that form the bases for our framework as they pertain to distributed groupware systems and CSCL. Given these requirements the multiple methods and data types afforded and constrained by the framework are described. Of particular significance are the multi-user distributed conditions that have forced us to develop new tools and use a diverse range of existing methods and techniques in novel ways to explore learning-technology relationships. Lastly, we describe some of the methodological perplexities facing researchers who combine multiple methods and data and some solutions for managing the data collection, analysis, integration, and interpretation process.

Learning in Networked Communities (LiNC)

The evaluation work described here is part of an interdisciplinary educational technology project called Learning in Networked Communities (LiNC). Two middle school physical science and two high school physics classes have been using educational groupware to plan, coordinate, carry out, and analyze science experiments coordinated between schools through computer networking. Student project groups can cut across age, across school, and across urban-rural areas. To support such activities we have developed a unique set of tools.

A single interface integrates a set of groupware tools with various collaboration and synchronous and asynchronous communication mechanisms (Koenemann, Carroll, Shaffer, Rosson, and Abrams, 1998). A Java-based networked learning environment called the "Virtual School" includes a collaborative science notebook that allows personal or shared workspaces for planning, developing, organizing, shared writing, and annotation of science projects. Communication tools built into the Virtual School include structured Web-based discussion forums, e-mail, real-time chat, video conferencing, and shared whiteboards. Application sharing is also directly integrated, allowing physically distributed students to remotely collaborate over and jointly control any single-user software that supports their activities in the Virtual School (e.g., sharing numerical data in spreadsheets). A server coordinates and preserves content centrally across users. The software and activities described above drive several requirements that structure our evaluation framework.

Requirements for Distributed Multi-user and CSCL Evaluation

Contextual evaluation

Evaluation methodologies for single-user computer interfaces have serious deficiencies when applied to the recent development of groupware interfaces (Greenberg, 1991), and consequently, there is no accepted mainstream approach to multi-user assessment. It is difficult to create controlled situations in the lab that reflect the social, motivational, organizational, and political dynamics so essential for groupware success found in actual work contexts (Grudin, 1990). Learning and technology integration cannot be divorced from contexts of use either. Subtleties of studentsí learning mediated by technology, teachers implementation of such technology constrained by real curriculum demands, and common classroom limitations such as insufficient time, too many students, and not enough computers make it imperative that CSCL technologies be implemented and subsequently evaluated in genuine classrooms. To adequately address these issues, contextual evaluation (Whiteside, Bennett, and Holzblatt, 1988; Holtzblatt and Jones, 1993) is well suited for assessing outcomes of technology use in the classroom. The necessity for framing research questions in authentic contexts has caused us to consider curriculum issues, longitudinal consequences, and multiple data collection methods as important contingencies surrounding evaluation.

Curriculum, longitudinal evaluation, and action research

Technology cannot be designed and evaluated in isolation of designing and evaluating curriculum and pedagogy. Focusing on curriculum during evaluation has become a greater concern in the development of education technology for two reasons: 1) The number and types of technology available in the classroom have significantly increased, and 2) there have been radical shifts in fundamental beliefs surrounding the goals and strategies to education (Roblyer, Edwards, Havriluk, 1997). Combining constructivismóa pedagogical shift from traditional instructionóand technology is a recipe for their co-development: When technology is introduced, curriculum and classroom practice change. Assuming teachers and curriculum will inherently develop and adapt to technological implementations can be disastrous. For educational technology to be successful, its design must actively and pervasively involve the co-design of curriculum and classroom practice.

Technology use in the classroom and learning inescapably emerge over time, often over weeks or months. To understand how learning unfolds in this context, extended periods of time are needed to understand how learning is affected by technology. This is especially the case when a wide range of issues is being considered in the evaluation or when evaluation has components of exploration. As a result, CSCL evaluation should be longitudinal to the extent possible.

Because technology, curriculum, and classroom practice co-evolve through deliberate design, and because we structure research questions in authentic classrooms (fieldwork), we frame our evaluation in action researchóa paradigm integrating traditional scientific objectives of analysis and explanation with the engineering objective of melioration. It is an iterative approach to combining theory and inquiry with practice by implementing change and reflecting on consequences (Lau, 1997). Researchers and practitioners (teachers) collaboratively carry out the evaluation, albeit at different levels. Approaching evaluation from this framework has been a natural incorporation into our entire *participatory design* life cycle that fully involves our teacher-collaborators (see Chin, Rosson, and Carroll, 1998).

Evaluation Framework for Distributed Multi-user Systems

Synergistic methodological pluralism

Based on the evaluation requirements described above, we have developed a multi-faceted evaluation framework that captures complex, distributed activities by combining quantitative human performance data and approaches under a qualitative research framework based on fieldwork. We have adopted a relatively new but increasingly promising pragmatist research paradigm (Tashakkori and Teddlie, 1998). This strategy uses a mixed-model research design, which fundamentally incorporates in all phases of the research process both quantitative and qualitative philosophies and approaches. Mixing methods in our research goes beyond purely methodological considerations. We have followed Brewer and Hunterís (1989) approach to including both quantitative and qualitative paradigms in all stages of the research process: problem formation,

measurement, building and testing theory, data collection/analysis, sampling, and the reporting of findings.

Mixed-model evaluation process

Our emphasis up to this point in our research has been on formative evaluation (Scriven, 1967)órich descriptions of classroom practice incorporating technology used to guide refinement and re-design. Figure 1 is a visual representation of our multi-faced evaluation framework. Quantitative deductive philosophies and qualitative inductive philosophies are applied to all levels of both quantitative and qualitative methods and data. The methods include interviews, surveys, questionnaires, focus groups, direct observation and field notes, contextual inquiry, video, system logs, and student constructed artifacts. We have had to substantially modify existing single-user methods and create new procedures for mechanically and analytically bridging methods and data. Central to these procedures are the integrated event scripts, critical incidents, and coding methods used to integrate the evaluation processes.

Multiple Evaluation Methods and Research Procedures

Surveys and interviews

We collect self-report measures from surveys and interviews, including close-ended questions about demographics, computer experience, and history associated with group learning. Much of this information is exploratory in nature. Likert-type rating scales are used to gather studentsí
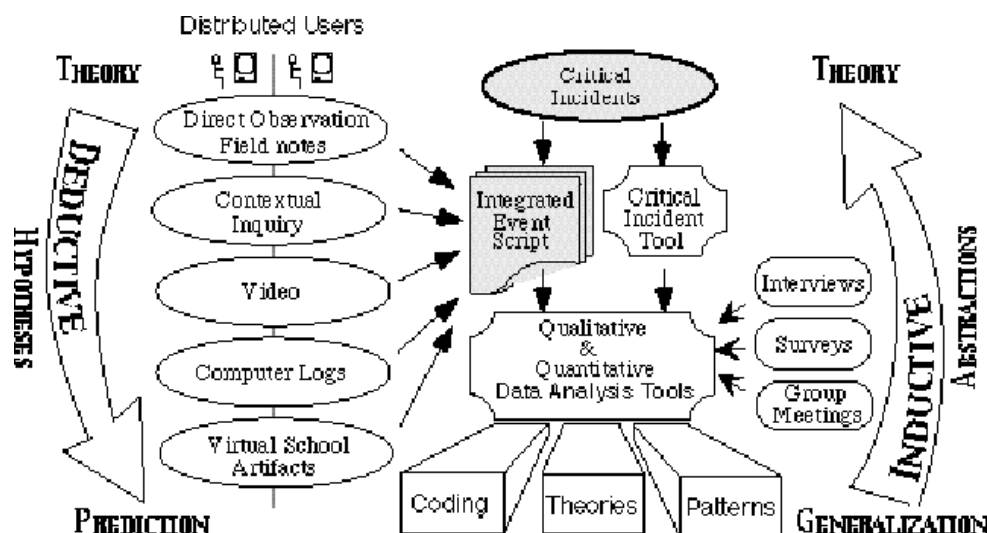

Figure 1. Distributed multi-user evaluation framework

preferences and attitudes toward science, collaboration, self-efficacy, and group learning. Many of these questions are constructed to answer *a priori* hypotheses based on the literature and theory about technologies impact on learning. Measures feed the formative evaluation, but they also serve as indicators of final outcomes resulting from pre-test to post-test comparisons following technological interventions each year. Because we cannot randomly assign students to classroom or project groups, the comparisons are based on quasi-experimental designs. To a lessor degree we have also administered teacher questionnaires with close-ended quantitative questions and open-ended qualitative questions. We also extensively interview teachers and students throughout the year. Although they are entirely qualitative, questions on the more structured side of the continuum are based on deductive claims from the literature, theory, and ongoing findings. This approach to questionnaires and interviewing lends itself to the quantitative and qualitative paradigms (Tashakkori and Teddlie, 1998).

Observational methods

We systematically sample field settings to carry out direct observations of collaboration and system use (Bogdan and Biklen, 1998; Coffey and Atkinson, 1996; Cohen and Manion; and Miles and Huberman, 1994). For any project group observation, we capture behaviors from both of the proximal groups, one at each school. Descriptive field notes (detailed objective descriptions of accounts) and reflective field notes (subjective ideas, reflections, patterns, and themes developed by the researchers) are meticulously recorded. Unlike the stratified sampling used to meet statistical representativeness for quantitative questionnaire data, *purposive* sampling is used for the qualitative methods because random sampling will often bias the representativeness of small samples (Miles and Huberman, 1994). We use maximum variation sampling as inductive and exploratory, and our theory driven sampling tends to be prediction and hypotheses based.

Contextual inquiry

During field evaluation, contextual inquiry (Holtzblatt and Jones, 1993; and Beyer and Holtzblatt, 1998) is used to collect rich descriptions of peoplesí activities in the context of their actual work. Contextual inquiry is a method used to inform engineering design by allowing people to describe their experiences while they work and learn. Some of this questioning is inductive, exploiting information gathering that is relevant to the immediate context. However, much of the questioning is determined a priori based on ongoing analysis of the data and established theoretical issues.

Video

During our observations, video and audio recordings capture behaviors. Using video as a research technique provides *thick descriptions* of behavior and activities. Video is particularly useful in our research because we capture co-discovery learning (Kennedy, 1989) when students verbalize their problem solving while they work and learn together. In addition to a long-shot camera view of students, scan converted images of screen interactions are captured and recorded to videotape. Video is suited to top-down planned

comparisons and bottom-up participant observations, encompassing both quantitative and qualitative methods and paradigms (Roschelle, in press).

Computer logs

Our software is instrumented for logging all computer interactions. We have developed complex routines for event capturing, and separate data filtering programs translate and aggregate system events generated by users into meaningful portrayals of usersí behavior with the system. The data is all time stamped for an accurate reconstruction of distributed system interactions. This information is particularly useful because it collects data from all the system users under different circumstances. This method furnishes an enormous amount of data that is subjected to statistical analysis. In addition to the quantitative use of logging data, complete transcripts of the information (artifacts) generated by students and the communication surrounding it are subjected to content analysis. Patterns are studied to confirm hypothesis, and new relationships are identified inductively.

Mechanical , Methodological, and Analytical Data Bridging

Dealing with the masses of data generated by the methods described above is an extremely daunting task. The variety of data collected from different researchers also makes it difficult to store, merge, share and retrieve data. An important first step is to develop a coding scheme that represents all the relevant factors to the data itself and all other types of data and documents that might be related to it. Trying to model all the relevant factors and the relationships between them early will facilitate developing a coding system for bridging data across the methods described below.

Critical incidents

One of the toughest decisions facing researchers using multiple methods with large amounts of data is to decide on what should be analyzed and when. Regardless of the approaches used, the research process must have a focus. One of the primary methods for determining our data analysis focus and for bridging analyses across methods is to collect and interpret userís *critical incidents* (Flannagan, 1954; and Meister, 1985). These are particularly good or bad experiences users have while trying to carry out tasks with our software systems. The critical incident technique was originally adapted to computer systems and construed as an "…interaction with a feature or element of the system that proved to be particularly easy or difficult, such that it greatly aided or hindered the userís progress in performing a task." (del Galdo, Williges, Williges, and Wixon, 1987). Critical incidents are typically gathered from a single user in a lab context. Because we study groups in field contexts with a much broader range of behaviors possible, in addition to software usability we have expanded the notion of a critical incident to include collaboration, communication, learning, and teacher practices.

We identify, collect, and analyze these incidents throughout the data collection process. Carroll, Koenemann-Belliveau, Rosson, and Singley (1993) have shown that the causes of critical incidents are not always immediate to their occurrence because they are temporally removed. The underlying causes also can be distributed across situations and

interactions with the system. Sets of causally related remote and distributed user interactions with the system that lead to a critical incident are called critical threads. Collections of these critical threads make up major usability themes. To date, identifying critical threads that make up critical incidents has been our most successful method of focusing the analysis and bridging major issues across methods and data types. Even significant quantitative findings from questionnaire data can contribute to critical threads. Having a variety of data types offers a richer source of information to develop critical threads, and the critical incidents/critical threads techniques are well suited to deductive and inductive approaches.

Formative evaluation of our software using a variety of researcher-driven methods has generated a large number of critical incidents. In order to enhance our understanding of these events and more fully involve those for who the events have significance, we have created and implemented a *collaborative critical incident tool*. This is a Web-based collaborative forum for posting, commenting on, elaborating, and developing critical incidents. All stakeholders in the softwareóteachers, developers, researchersóexpress incident criticality quantitatively and contribute to the understanding of what precipitated an incident, what underlying causes contributed to it, and what the past and future implications are as a result of its occurrence. This approach to evaluation has expedited and broadened our understanding of major themes of significance by further expanding on critical threads. Not only has it facilitate discussion among researchers and developers, but another important consequence of using this method has been allowing teachers to contribute their own critical incidents, providing critical thread information that is often inaccessible to evaluators. By sharing our interpretations with those we study (teachers, students, and community members) and incorporating their shared understanding, we have provided vital validity to evaluation findings. We have found this approach to be the most successful form of respondent validation to evaluation outcomes. Much of the discussion leads to scenarios for supporting and enhancing good experiences, and mediating or eliminating bad ones.

Integrated event scripts

Transforming data into a state that permits us to maximize methodological and analytical data bridging is difficult with the multiple forms of data. Compounding the problem considerably is developing an approach to the difficult empirical situation of analyzing multiple physical locations (sometimes as many as three). We have constructed an approach involving the synchronization and subsequent interleaving of observer notes and data collected from the connected sites. It is an elaborate scheme for organizing and collating the many forms of data being collected about student learning (real-time contextual interviews, videotaped sessions, field notes, computer logs, student artifacts, and critical incidents) (Figure 1). In this section we will describe how we create *integrated event scripts*-synchronized, episodic, multi-threaded documents that capture different user and group activities from multiple data in different contexts. Once researcher pairs (capturing two sides of a distributed group interaction) return from classrooms with field notes, contextual inquiry results, and video; each researcher records, combines, and elaborates on their proximal end of the group interaction.

Video protocol analysis is used immediately following the observation to transcribe behaviors using CVideo (Roschelle, 1996). Field notes are directly incorporated into the transcript. Reflective field notes are distinguished from descriptive field notes with notation. Contextual inquiry data and critical incidents are also represented with notation. All video recordings are time stamped to match the server clock that time stamps the computer logs. Transcripts of the video, field notes, and contextual inquiry are linked to the video using the time from the videotape matching the server. The Cvideo software is used in future analyses to automatically search and locate video episodes time stamped in the transcripts. Once the transcription is complete, the computer logs are integrated temporally in the appropriate locations, and the transcripts from the proximal groups are combined into a single document. A different type font is used to distinguish between the proximal group transcripts, and a third font is used to denote the computer logs. Researchers return to the integrated event script individually and further expand on their observer comments.

We have a number of video editing and mixing systems that allow us to combine video from different sources. Split screen video (x4) and audio mixing are used to reconstruct synchronous events that occur at multiple school sites. The four video sources (long-shots and screen captures from each subgroup) are combined onto a single tape. A mixer is used to split the audio signals onto the left and right channels. This allows us to listen to a single side (school) of the group interaction if needed during repeated viewing. The mixing of audio onto a single channel can make it imperceptible because distributed groups are often performing local work that is not coordinated with remote partners, producing incoherent, disjunct conversations. All of the video has to be synched within 1 or 2 seconds of each other; otherwise, the synchronization of behaviors and events are difficult to follow.

Once the integrated event script has been created, the video and audio mixed, and the researchers have had a chance to study the distributed interaction from the variety of data, the researchers sit down in pairs with the event script and watch the video together. Repeated viewing is often necessary to accurately and thoroughly reconstruct distributed events. From researcher collaboration a considerable amount of understanding develops regarding usersí communication and collaboration. Only from understanding what happened at the remote site can a researcher understand what happened at the proximal group studied. This information further feeds into the integrated event scripts as observer interaction comments. Temporally synching multiple data that has been collected from distributed research settings and forging sustained, collaborative work over this data is imperative to understanding how students learn using groupware.

Coding quantitative and qualitative data

Although we have described several unique procedures for data bridging, beyond these methods entire data sets need to be aggregated. By combining the integrated event scripts with collaborative critical incidents, interviews, and surveys, the total data set can be analyzed as a whole. This is accomplished by continuously assigning codes to the data (Coffey and Atkinson, 1996; and Bogdan and Biklen, 1998). Codes are key words or phrases that represent the conceptual issues under study. Coding is the process of

systematically organizing the data, reducing it into manageable units, synthesizing the units, detecting patterns, and presenting the information for interpretation. QSR NUD*IST, a qualitative data-analysis software package (Scolari, 1997), is used to assign tags or labels from categories and concepts (coding chunksówords, sentences, paragraphs, documents, video, previous findings). The categories and concepts are established a priori from theories and literature (Miles and Huberman, 1994) or emerge bottom-up from the analysis (Glaser and Strauss, 1967; and Strauss and Corbin, 1994). Codes have hierarchical relationships, but they can also be orthogonal or overlapping. NUD*IST supports indexing, searching, and theory building by allowing a range of data from different methods to be grouped according to a coding category or group of categories. The coding process results from early and continued focus on empirical findings and theory, and intense, sustained collaborative interactions between researchers. Analyzing the data, therefore, becomes a shared development and learning process for the research team (Nyerges, Moore, Montejano, and Compton, 1998).

Bringing data fragments together through category linkages allows summative conclusions to be drawn about the overall effectiveness of learning, technology, and the curriculum and classroom practice framing them. The data is examined to yield a final analysis composed of both quantitative (e.g., an increase in cooperative planning) and qualitative (e.g., a description of how the increased planning was accomplished) measures of whether and how the Virtual School activities have had an impact on studentsí learning experiences. Aggregating, mapping, and measuring code incidence is used as a quasi-quantitative form of content analysis to base summative conclusions.

One of our main objectives is to analytically combine strategies from quantitative and qualitative analyses by taking data from one tradition and converting it into another tradition. This grants the use of alternative techniques for analyzing the same data. As Miles and Huberman (1994) suggest, we "quantize" the qualitative data where appropriate. Analytic significance is obtained by analyzing event counts (frequency), duration of behaviors (time), and errors encountered (accuracy). Qualitative information converted into numerical codes can then subjected to statistical analysis once inter-coder reliability is determined. We also have plans to convert some of the data into ranks or scales based on expert reviews. We also "qualitize" quantitative data by converting it to qualitative data. For example, descriptive statistics from survey data or significant inferential findings are taken into the qualitative analysis tool and coded along with other qualitative data. These approaches to combining analyses bridge data to other methods and data, and bridge procedures and philosophies of quantitative and qualitative paradigms.

Discussion and Conclusions

Several important conceptual factors contributed to our choice of a mixed framework approach. In a review of the literature Greene, Caracelli, and Graham (1989) identified 5 purposes for using mixed-method designs: triangulation, complementary, developmental, initiation, and expansion. In Table 1 we list the 5 reasons corresponding to Greene et. al.

for our decision to use mixed methods in order of importance, with the rationale for their ranking.

Table 1. Conceptual basis for our mixed-model design.

Ranking	Purpose	Rationale
1	Complementary	Our primary decision to mix methods is to further the interpretability and meaningfulness of our findings by capitalizing on the strengths of multiple methods and thus minimizing biases across methods.
2	Expansion	We expand our inquiry by choosing multiple methods and apply them to different areas of the study. Certain methods are best applied to different evaluation questions.
3	Triangulation	We increase validity and reduce sources of variance by converging findings across methods. Method, data, investigator, and theory biases are limited.
4	Initiation	While we don't seek out contradiction, when it does arise, we capitalize on exploring it from different perspectives.
5	Developmental	Lastly, although we do not specifically use results from one method to develop another, we do use all of our findings to continually refine methods in an iterative fashion.

All methods have certain advantages, and all methods have inherent weaknesses. We have capitalized on strengths and minimized limitations by using multiple methods that have patterned diversity (McGrath, 1994). Essential to this approach is the triangulation techniques (pioneered by Campbell and Fiske, 1959) that bridge data and process in an iterative cycle. Our research incorporates all four classic triangulation methods described by Denzin (1978), including data, investigator, theory, and methodological triangulation. We also incorporate time triangulation, space triangulation, and combined levels (individual, group, school) triangulation. Triangulation is desirable when complex phenomenon are being studied and when more holistic approaches are needed to explain the phenomenon. Three other important conditions particularly warrant triangulation in education: 1) when different teaching methods are being evaluated, 2) when controversial issues of education are being evaluated, and 3) when approaches to research are not established or yield a limited view of the topic under study (Cohen and Manion, 1994). Our research is characterized by all these factors.

Unlike what is typically presented in scholarly journals characterizing research as sequential and objective, our approach more fully accounts for how "real research is often confusing, messy, intensely frustrating, and fundamentally nonlinear" (Marshall and Rossman, 1999). Furthermore, the mixed-model design genuinely reflects the process of going between inductive and deductive models of thinking in research (Creswell, 1995). This paper has provided some guidelines lacking in the research literature for systematically integrating eclectic methods and data. Using different data from different sources for triangulation will not necessarily produce a more authentic and holistic account of the social world. However, the juxtaposition of methods and data through integrated event scripts and critical incidents significantly reduces the complexity of researchersí understanding when evaluating distributed systems from multiple perspectives.

Acknowledgements

Bibliography

Avison, D., Lau, F., Myers, M., and Nielsen, P. A. (1999). Action Research. *Communications of the ACM, 42*(1), 94-97.

Beyer, H., and Holtzblatt, K. (1998). *Contextual design: Defining Customer-centered systems*. San Francisco, CA: Morgan Kaufmann.

Bogdan, R. C., and Biklen, S. K. (1998). *Qualitative research for education: An introduction to theory and methods*. Boston: Allyn and Bacon.

Brewer, J., and Hunter, A. (1989). *Multimethod research: A synthesis of styles*. Newbury Park, CA: Sage.

Campbell, D., and Fiske, D. W. (1959). Convergent and discriminate validation by the multitrait-multimethod matrix. *Psychological Bulletin, 4*, 297-312.

Carroll, J. M., Koenemann-Belliveau, J., Rosson, M. B., and Singley, M. K. (1993). Critical incidents and critical themes in empirical usability evaluation. In *People and Computers VIII. Proceedings of the HCI'93 Conference* (pp. 279-292). Cambridge, England: Cambridge University Press.

Carroll, J. M., and Rosson, M. B. (1995). Managing evaluation goals for training. *Communications of the ACM, 38*(7), 40-48.

Chin, G., Rosson, M. B., and Carroll, J. M. (1997). Participatory analysis: Shared development of requirements from scenarios. In *Proceedings of Human Factors in Computing Systems, CHI'97 Conference* (pp. 162-169). New York: ACM.

Coffey, A., and Atkinson, P. (1996). *Making sense of qualitative data: Complementary research strategies*. Thousand Oaks, CA: Sage Publications.

Cohen, L., and Manion, L. (1994). *Research methods in education*. London: Routledge.

Cook, T. D., and Campbell, D. T. (1979). *Design and analysis of quasi-experiments for field settings*. Chicago: Rand-McNally.

Creswell, J. W. (1994). *Research design: Qualitative & quantitative approaches*. Thousand Oaks, CA: Sage.

del Galdo, E. M., Williges, R. C., Williges, B. H., and Wixon, D. R. (1987). A critical incident evaluation tool for software documentation. In L. S. Mark, J. S. Warm, and R. L. Huston (Eds.), *Ergonomics and Human Factors* (pp. 253-258). New York, NY: Springer-Berlag.

Denzin, N. K. (1970). *The research act in sociology: A theoretical introduction to sociological method*. London: the Butterworth Group.

Denzin, N. K. (1978). *The research act: A theoretical introduction to sociological methods (2nd ed.)*. New York: McGraw-Hill.

Flanagan, J. C. (1954). The critical incident technique. *Psychological bulletin, 51*(4), 327-358.

Gallivan, M. J. (1997). Value in triangulation: A comparison of two approaches for combining qualitative and quantitative methods. In A. S. Lee, J. Liebenau, and J. I. DeGross (Eds.), *Information systems and qualitative research* (pp. 417-443). London: Chapman & Hall.

Glaser, B., and Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative research*. Chicago, IL: Aldine.

Greenberg, S. (Ed.). (1991). *Computer supported cooperative work and groupware*: Academic Press.

Greene, J. C., Caracelli, V. J., and Graham, W. F. (1989). Toward a conceptual framework for mixed-method evaluation designs. *Educational Evaluation and Policy Analysis, 11*, 255-274.

Grudin, J. (1990). Groupware and cooperative work: Problems and prospects. In B. Laural (Ed.), *The Art of Human Computer Interface Design* . Reading, MA: Addison-Wesley.

Holtzblatt, K., and Jones, S. (1993). Contextual inquiry: A participatory technique for system design. In D. Schuler and A. Namioka (Eds.), *Participatory design: Principles and practices* . Mahwah, N. J.: Lawrence-Erlbaum.

Kay, J., and Thomas, R. C. (1995). Studying long-term system use. *Communications of the ACM, 38*(7), 61-69.

Kennedy, S. (1989). Using video in the BNR usability lab. *ACM SIGCHI Bulletin, 21*(2), 92-95.

Koenemann, J., Carroll, J. M., Shaffer, C. A., Rosson, M. B., and Abrams, M. (1998). Designing collaborative applications for classroom use: The LiNC Project. In A. Druin

(Ed.), *The design of children's technology* (pp. 99-122). San Francisco: Morgan-Kaufmann.

Lau, F. (1997). A Review on the use of action research in information systems studies. In A. S. Lee, J. Liebenau, and J. I. DeGross (Eds.), *Information Systems and Qualitative Research*. London: Chapman and Hall.

Marshall, C., and Rossman, G. B. (1999). *Designing qualitative research, 3rd Edition*. Thousand Oaks, CA: Sage.

McGrath, J. E. (1994). Methodology matters: Doing research in the behavioral and social sciences.

Miester, D. (1985). *Behavioral Analysis and Measurement Methods*. New York: Wiley.

Miles, M. B., and Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook*. Thousand Oaks, CA: Sage.

Nielson, J. (1993). *Usability Engineering*. Boston: Academic Press.

Nyerges, T., Moore, T. J., Montejano, R., and Compton, M. (1998). Developing and using interaction coding systems for studying groupware use. *Human-Computer Interaction, 13*, 127-165.

QSR, NUD*IST (1997). *Software for qualitative data analysis*. Thousand Oaks, CA: Scolari.

Roblyer, M. D., Edwards, J., and Havriluk, M. A. (1997). *Integrating educational technology into teaching*. Upper Saddle River, N. J.: Prentice Hall.

Roschelle, J. (1996). Video annotation and analysis tool [On-line]. Available: http://www.slip.net/~jeremy/CVideo/index.html.

Roschelle, J. (in press). Choosing and using video equipment for data collection. In A. Kelly and R. Lesh (Eds.), *Research Design in Mathematics & Science Education*. Amsterdam: Kleuwer.

Ruhleder, K., and Jordan, B. (1997). Capturing complex, distributed activities: Video-based interaction analysis as a component of workplace ethnography. In A. S. Lee, J. Liebenau, and J. I. D. Gross (Eds.), *Information Systems and Qualitative Research*. London, UK: Chapman and Hall.

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

Scriven, M. (1967). The methodology of evaluation. In R. Tyler, R. Gagne, and M. Scriven (Eds.), *Perspectives of curriculum evaluation* (pp. 39-83). Chicago: Rand McNally.

Strauss, A., and Corbin, J. (1994). Grounded theory methodology: An overview. In N. K. Denzin and Y. S. Lincoln (Eds.), *Handbook of qualitative research*. Thousand Oaks, CA: Sage.

Tashakkori, A., and Teddlie, C. (1998). *Mixed methodology: Combining qualitative and quantitative approaches*. Thousand Oaks, CA: SAGE.

Whiteside, J., Bennett, J., and Holtzblatt, K. (1988). Usability engineering: Our experience and evolutions. In M. Helander (Ed.), *Handbook of Human-Computer Interaction* (pp. 791-817). North Holland: Elsevier Science.

Authorsí addresses

Dennis C. Neale (dneale@vt.edu)
Center for Human-Computer Interaction; Department of Computer Science;
Virginia Tech 660 McBryde Hall; Blacksburg, VA 24061-0106. Tel. (540) 231-6931.
John M. Carroll (mailto:carroll@vtopus.cs.vt.edu)
Center for Human-Computer Interaction; Department of Computer Science;
Virginia Tech 660 McBryde Hall; Blacksburg, VA 24061-0106. Tel. (540) 231-6931.