Strategies for Evaluating Information Visualization Tools: Multi-dimensional In-depth Long-term Case Studies

Ben Shneiderman[#]*, Catherine Plaisant[#]
[#]Human-Computer Interaction Laboratory, Institute for Advanced Computer Studies and
*Computer Science Department
University of Maryland
{ben, plaisant}@cs.umd.edu

ABSTRACT

After an historical review of evaluation methods, we describe an emerging research method called Multi-dimensional In-depth Long-term Case studies (MILCs) which seems well adapted to study the creative activities that users of information visualization systems engage in. We propose that the efficacy of tools can be assessed by documenting 1) usage (observations, interviews, surveys, logging etc.) and 2) expert users' success in achieving their professional goals. We summarize lessons from related ethnography methods used in HCI and provide guidelines for conducting MILCs for information visualization. We suggest ways to refine the methods for MILCs in modest sized projects and then envision ambitious projects with 3-10 researchers working over 1-3 years to understand individual and organizational use of information visualization by domain experts working at the frontiers of knowledge in their fields.

Categories and Subject Descriptors

H.5.2 [**Information Interfaces and Presentation**]: User Interfaces – *Graphical user interfaces (GUI), Interaction styles, Screen design, Evaluation/methodology*.

General Terms

Experimentation, Human Factors.

1. INTRODUCTION

The goals of human-computer interaction (HCI) evaluation have been shifting to accommodate the rising aspirations of interface designers and HCI researchers. The pendulum of scientific research is once again swinging from the height of reductionist thinking that emphasizes tight laboratory control towards the situated strategies that emphasize ethnographically-oriented and longitudinal participant observation. We seek to encourage information visualization researchers to study users doing their own work in the process of achieving their goals. An emerging research method called Multi-dimensional In-depth Long-term Case studies (MILCs) seem well adapted to study the creative activities that users of information visualization systems engage in [26].

BELIV 2006 Venice, Italy.

In the term "Multi-dimensional In-depth Long-term Case studies" the *multi-dimensional* aspect refers to using observations, interviews, surveys, as well as automated logging to assess user performance and interface efficacy and utility. The *in-depth* aspect is the intense engagement of the researchers with the expert users to the point of becoming a partner or assistant. *Long-term* refers to longitudinal studies that begin with training in use of a specific tool through proficient usage that leads to strategy changes for the expert users. *Case studies* refers to the detailed reporting about a small number of individuals working on their own problems, in their normal environment.

Longitudinal studies have been carried out in HCI and in some information visualization projects, but we propose to refine the methods and expand their scope. The controversial question is how far information visualization researchers can go in measuring the utility of their tools by the success achieved by the users they are studying.

2. HISTORICAL REVIEW OF EVALUATION METHODS

In the 400 years since Francis Bacon (1561-1626) first promoted reductionist thinking, scientific research was closely linked with controlled experiments. The strategy was for researchers to vary a small number of independent variables among a small number of treatments to determine the impact on a small number of dependent variables. All other factors were to be kept constant to avoid bias.

For example, physicists, starting with the apocryphal story of Galileo (1564-1642), varied the length of string on a pendulum from 40 to 50 to 60 centimeters (first independent variable with three treatments) while changing the weight of the pendulum from 1 to 2 kilograms (second independent variable with two treatments). The room temperature, thickness of string, altitude above the ground, and initial displacement might all be kept constant so as to minimize the impact of these potentially biasing effects. The goal would be to study the impact of changing the independent variables on the time for each pendulum swing, the dependent variable. The goal was to understand fundamental principles that would be generalizable to many pendulums (theory), and maybe even influence the design of clocks (practical problem), or ultimately improve the accuracy of timekeeping (broader goal).

Many generations of physicists, chemists, and other scientists successfully applied these reductionist strategies of scientific research, but laboratory studies often became ever more distant from practical problems and broader goals. Physicists went down the road of developing high-powered synchrotrons to produce extreme conditions that never occur in the natural world, sometimes producing fascinating discoveries, but sometimes diverging from solving practical problems and only occasionally advancing broader goals.

In emerging scientific fields, such as agricultural biology, statisticians such as Ronald Fisher (1890-1962), extended the notions of controlled experimentation to support testing of farming strategies, even when controls for rainfall, sunlight, or soil conditions could not be precisely maintained. Perceptual and motor skill psychologists soon adopted Fisher's methods to studying human performance. They wanted to measure how well humans could hear sounds, see images, or point at targets while reducing the individual differences that so persistently plague such studies. Cognitive psychologists followed their lead and defined narrow tasks from which they hoped to form generalizable theories of human problem solving and decision making.

By the 1970s human factors researchers were reshaping these methods to understand human performance so as to refine design of technologies such as air traffic control systems, airplane cockpits, or automobile dashboards. During the 1980s controlled experimental methods were applied to user interfaces, first for devices such as keyboards and the mouse, and later for interface issues such as command consistency and menu design.

As commercial and consumer applications expanded, desktop, mobile device and web site developers needed to make so many design decisions that novel testing methods became necessary. The archives of research-oriented controlled experiments were still useful for guidance, but usability testing became the norm for product developers[4][8]. Three to ten users would be given a set of typical tasks to see where they ran into trouble. The outcome of such testing was a report that identified common problems, possibly ranked by the difficulty in making revisions. Then changes would be made and a new usability test conducted, possibly within days or weeks. This iterative process led to rapid refinement of interfaces and a better understanding of the role of layout, terminology, and consistency, plus increased awareness of task frequencies and sequences.

These laboratory-like usability tests, even without control treatments, were criticized by some analysts who pointed out the situated nature of most work activity and consumer uses. They pushed for more field or case study styles of tests in the users' workplace, where interruptions, supplementary support materials, and social exchanges were common. In addition, a central change was the shift from supplying a standard set of tasks to enabling usability test participants to carry out their normal work or personal tasks.

A further development was the adoption of action research approaches that had been pioneered by Lewin [19]. In action research, the case study method went further, involving the researcher more explicitly in the work being studied [34]. The key driver for action research is to find ways to change, specifically to improve, processes over and above the search for knowledge about them. A good discussion of the issues is included in the Philosophical and Methodological chapter of Argyris's book, *Action Science* [2].

Those devoted to field studies adopted ethnographic methods with in-depth study of a small number of subjects carrying out their normal work using the new interfaces [10][32]. These ethnographic and sometimes longitudinal studies made for more realistic tests, but undermined the goal of carefully controlled conditions; a tradeoff that advocates were happy to make. Using a web search on HCI literature (www.hcibib.org) references to 'ethnographic' grew from 8 in the 1980s to 90 in the 1990s to 118 in the first six years of the 2000s, while references to 'longitudinal' grew from 17 to 67 then stabilized at 35 for the first six years of the 2000s. Case study research methods differ from ethnographic studies by establishing in advance the plausible rival hypotheses [35]. Case study researchers collect evidence to support these rival hypotheses such as that a new tool is beneficial. These field and case study styles of research, often infused by the ethnographic notions of participant observation, laid the basis for the emergence of a new paradigm for human-computer interaction evaluation.

3. THE CHALLENGE OF INFORMATION VISUALIZATION EVALUATION

As the field of information visualization matures, the tools developed in our research laboratories are reaching users. The reports of usability studies and controlled experiments are helpful but there is a growing desire for alternative methods of evaluation in order to present actionable evidence of measurable benefits that will encourage more widespread adoption of visualization [23]. Information visualization is usually part of some creative activity that requires users to make hypotheses, look for patterns and exceptions, and then refine their hypothesis. Users might find surprising results that shake their established beliefs, provoke new insights, and possibly lead to important discoveries. Users often need to look at the same data from different perspectives and over a long time. They may need a variety of tools to achieve their goals, repetitively exporting and importing data. Users are likely to collaborate with other users and they may be able to formulate and answer questions they didn't anticipate having before looking at the visualization. Finally, discoveries can have a huge impact but they occur very rarely, making it difficult – if not impossible – for someone to be observing when a discovery occurs.

In the remainder of the paper, we look at how Multi-dimensional In-depth Long-term Case studies (MILCs), which were developed to evaluate creativity support tools, might be used to evaluate information visualization tools.

4. A NEW PARADIGM FOR EVALUATION

Multi-dimensional In-depth Long-term Case studies (MILCs) have been embraced by the small but growing community of researchers who are studying creativity support tools [28]. The users of information visualization and creativity support tools are generally knowledgeable domain experts carrying out leading work in their fields: software engineers, architects, molecular biologists, lawyers, physicians, etc. They are working on challenging problems that are at the limit of what is known or common practice in their fields. They may be engaged in inventing, innovating or discovering. They want well-designed tools that are in harmony with basic human performance (perceptual, motor, cognitive, etc.), support rapid error-free performance of common tasks, and provide interfaces in which creative exploration is easy. These tools should support advanced services such as search, collaboration and dissemination, as well as flexible composition, hypothesis generation, and history keeping that enables rapid backtracking and macro-making [27].

The dilemma for those studying creativity support tools is how to evaluate and improve their effectiveness. Controlled experiments of specific features seem too narrow as do gross comparisons of one tool versus another. Controlling for individual differences seems nearly impossible and specifying tasks is somehow at odds with the goals of supporting innovation or discovery. Psychologists have tried to study creativity processes (e.g. random stimuli vs. guided discovery) with toy-like laboratory tasks such as asking subjects to describe as many uses of bricks as they can within 15-minutes; then the lists are assessed by review panels for quantity and quality [30]. Others have studied children's creativity with artistic tasks such as drawings of imaginary forests which were rated by a panel of artists [1]. Even with such constrained tasks the ratings can be controversial and experimental outcomes hard to discern.

In the area of information visualization efforts have been made to combine empirical studies with more naturalistic and creative situations. A rare example is the work of Saraiya et al. [25] who asked biology students to use several tools for a few hours, look at various datasets and report all the insights they gathered from this experience. The researchers were then able to classify the insights generated and compare the frequency and type of insights generated for the different tools. Much can be learnt from such experiments but the researchers commented that it was difficult to motivate the participants and that the insights reported were of modest interest. The 2003 and 2004 InfoVis contests [11] were also useful to informally compare tools as participants were asked to self report insights generated while using the tools over several months. Here obtaining an award was the motivation for some participants, but there was no quantitative analysis of the results. In contrast the Visual Analytics VAST contest [33] will combine both qualitative and quantitative metrics by using synthetic dataset which provides ground truth against which insights can be evaluated.

The new paradigm of MILCs builds on the notion of field or case studies using ethnographical participant observation methods, plus interviews, surveys, and automated logging of user activity. Promoters of MILCs for creativity support tools suggest long-term observations over weeks or months are necessary to fully understand how domain experts work and how they apply creativity support tools [28]. In our recent example [26], five users of an advanced data analysis and visualization tool were studied for up to 6 weeks each on a one hour per week basis. These expert users received training and agreed to participate in the study, but eventually two users dropped, leaving three participants (a biologist, statistician, and meteorologist) who gave detailed and constructive comments about interface features and problems. Participants were given additional help in using the tool and software improvements were made on request. This intense level of interaction was helpful in understanding the problems users were dealing with and ultimately enabling all three users to make important scientific discoveries (e.g. a strong association between a specific gene and body composition).

This project provides one model for MILCs. There are others from early work on iterative testing in developing an information retrieval system [15], recent research on collaboratories [22] as well as some studies of educational environments [20], designers [6] and new media artists [7][9]. In every case, the researchers prepared a guiding set of research questions to narrow their focus, then developed trust and rapport with the subjects. Early stages of the process require careful steps to gain entry, permission, and participation of subjects, followed by intense discussions which provide the key data for researchers. Later stages include discussions between the researchers and the expert users, and even comments from the users about the researchers's report [13][16].

Outcomes for MILCs are generally in two categories:

1) the refinement of the tool and an understanding of general principles or guidelines for the design of such tools.
2) the achievement of the expert users' goals, by way of their use of the tool.

Since researchers often become engaged with the expert users to help them achieve their goals, this approach may be troubling to those who believe in the need for ethnographers and anthropologists to resist interfering with the culture or community they are studying.

5. LESSONS FROM ETHNOGRAPHIC OBSERVATIONS IN HCI DESIGN

Since interface users form a unique culture, ethnographic methods for observing them have been used by researchers, mostly to guide the design of novel interfaces [17] or to improve existing ones [24]. Ethnographers join work or home environments to listen and observe carefully, sometimes stepping forward to ask questions and participate in activities [5][12][14][21]. As ethnographers, user-interface designers gain insight into individual behavior and the organizational context. User-interface designers differ from traditional ethnographers; in addition to understanding their subjects, user-interface designers focus on interfaces for the purpose of changing and improving those interfaces. Whereas traditional ethnographers immerse themselves in cultures for weeks or months, user-interface designers usually need to limit this process to a period of days or even hours, and still to obtain the relevant data needed to influence a redesign [18]. Ethnographic methods have been applied to office work [31], air-traffic control [3], and other domains.

Unfortunately, there are many ways in which ethnographic observation can go wrong: it is easy to misinterpret observations, to disrupt normal practice, and to overlook important events. Following a validated ethnographic process reduces the likelihood of these problems. Guidelines for preparing for the evaluation, performing the field study, analyzing the data, and reporting the findings might include the following [24]:

Preparation

- Understand organization policies and work culture.
- Familiarize yourself with the system and its history.
- Set initial goals and prepare questions.
- Gain access and permission to observe or interview.

Field Study

- Establish rapport with managers and users.
- Observe or interview users in their workplace, and collect subjective and objective quantitative and qualitative data.
- Follow any leads that emerge from the visits.
- Record your visits.

Analysis

- Compile the collected data in numerical, textual, and multimedia databases.
- Quantify data and compile statistics.
- Reduce and interpret the data.
- Refine the goals and the process used.

Reporting

- Consider multiple audiences and goals.
- Prepare a report and present the findings.

These notions seem obvious when stated but they require interpretation and attention in each situation. For example, understanding the differing perceptions that managers and users have about the efficacy of an interface might alert researchers to the varying frustrations that each group will have. Learning the technical language of the users is also vital for establishing rapport. Data collection can include a wide range of subjective impressions that are qualitative or of subjective reactions that are quantitative, such as rating scales or rankings. Objective data can consist of qualitative anecdotes or critical incidents that capture user experiences, or can be quantitative reports about, for example, the number of errors that occur during a one-hour observation of six users.

Deciding in advance what to capture is highly beneficial, but remaining alert to unexpected happenings is also valuable. Written report summaries have proved to be valuable, while raw transcripts of every conversation are too voluminous to be useful.

A well-designed ethnographic process has many benefits. It can increase trustworthiness and credibility, since designers learn about the complexities of an organization firsthand by visits to the workplace. Personal presence allows designers to develop working relationships with several end users to discuss design ideas. The guidelines from ethnographic observations reported in this section related to studies conducted as part of the design or redesign of an interface. In the next section we discuss guidelines for the use of applied ethnographic methods for evaluation and refinement of information visualizations.

6. CONDUCTING MULTI-DIMENSIONAL IN-DEPTH LONG-TERM CASE STUDIES OF INFORMATION VISUALIZATIONS

The ethnographic methods envisioned in the MILCs can be adapted to serve the goals of information visualization evaluation. While variations are possible, we focus on the situation in which a researcher has developed a novel approach to information visualization and seeks to evaluate its efficacy. This situation often emerges in academic research and in early stages of commercial product development. The researchers or developers are eager to find the strengths and weaknesses of their new information visualization tool. Their larger goals are to refine the tool and to claim enough success to warrant academic recognition (an accepted doctoral dissertation or scientific paper) or further commercial development. Still larger goals might be to understand aspects of human problem solving, processes of technology adoption, roadblocks to strategy revision, or social processes that are necessary for organizational success with new information visualization tools.

We begin with a proposal for modest MILCs that could be applied in existing projects and then suggest a much more ambitious application of MILCs that require major new funding.

Researchers or developers who wish to evaluate their information visualizations with the MILC approach can get started with modest effort by identifying 3-5 domain experts who are willing to cooperate for a period of several weeks to several months. The novel tool has to be developed and tested sufficiently with usability studies to remove obvious problems and ensure that the tool has a reasonable level of reliability and support for basic features (e.g. import-export, saving partial results, printing, logging, etc.). We summarize our proposed MILC guidelines as follows:

- Specify your research questions and goals, keeping them focused. Focus your attention on specific aspects of the tool and its use. When we studied HCE users [26], we concentrated on the rank-by-feature framework, with minimal attention to use of the dendrogram and parallel coordinates.

- Identify 3 to 5 users. Ideally their expertise and goals will be varied to provide different perspectives. There is no need to start with all users at once. A staggered start may help iron out the training and procedure. Expect that some users might drop out. Be flexible and helpful to users.

- Document the current method or tool being replaced or augmented by the new tool. Document the current version of the tool being tested, and record changes made to its design.

- Determine what would constitute professional success for the users. For scientists this might be the submission of a scientific paper. For biologists it might be the discovery of a new drug. Varying levels of success might be described as well, i.e. submission to different journals, or submissions only partially linked to the use of the tool may have different values. If the timeframe for major breakthroughs is too long then smaller steps toward success might have to be described.

- Establish a schedule of observation and interviews. At first visits should be long (a few hours) and regular (e.g. every day or every week). Later on, visits could be shorter (even possibly limited to phone interviews) and more spaced out, but researchers should always be ready to stay longer if some major breakthrough has been made and needs to be documented. Be flexible.

- Instrument the tool to record usage data e.g. features used, frequency of use, datasets opened or saved etc. Obtain appropriate permissions. Complement as needed by collecting screen shots, sample datasets or generated reports at each visit. Take photos of materials printed or drawn.

- Provide an attractive log book to users for recording comments, problems, and insights gathered. Encourage users to record difficulties and frustration, as well as successes. An attractive log book also serves as a reminder to users that they should be recording their experiences, something that many researchers are reluctant to do.

- Provide training. Observing how hard it is to learn a tool may be interesting, but the focus should be on bringing users to the level of expert users. Training will most likely continue over the initial visits and interviews. Having a specific person always available to answer questions is beneficial.

- Conduct visits and interviews. Establish personal contact with users, then ask them to reflect on their use of the tool and the insights generated. Inquire about other tools being used in conjunction with the tool being studied (including paper and pencil). Inquire about collaborations which take place. Discuss log book entries. Finally, reflect and summarize what the users have learned and how much progress they have made toward their goal. Write down your insights immediately to ensure that you record important details.
- Encourage users to continue using the best possible tool for the task, to avoid a situation where users try to please the researcher by using the new tool while another classic one would have been more appropriate.
- Modify tool as needed. When appropriate, the tool might have to be modified or extended to provide the functionality users need. In some cases the researcher might just assist users instead of building additional features (e.g. if the data format is not accepted by another tool the users need to use, it might be more effective to convert the specific file manually than building a generic conversion). Be flexible
- Document success and failures. Immediately after each visit or interview, reflect on lessons learned. Ask users to check the summaries or final reports you write.

As we conduct more of these modest MILCs to complement the traditional evaluation methods, we will refine our guidelines and report on the success and failures of the method itself.

We envision that much more ambitious applications of MILCs would have enormous benefits for researchers and developers. Scaling up by an order of magnitude in terms of number of users and duration of the observations is clearly beneficial. Observations of dozens of users over months and years would do much to improve the reliability, validity, and generalizability of the results. If dozens of software engineers using a source code static analysis visualization tool were found to eagerly adopt this new tool and increased their usage over several months this would be compelling evidence. If they found that their programs had fewer bugs, or were developed more rapidly, or became better candidates for parallelization that would give still greater support for the value of their new tool. Long-term studies over a year or more document the experiences of artists using technology in MILC-like studies [7].

An even more ambitious application of MILCs would be to study social creativity with teams and larger organizations. Many commercial information visualization tools have been adopted by organizations with hundreds or thousands of users in collaborative applications, but we have little understanding of what generates success in these environments. For example, Spotfire is used by hundreds of pharmaceutical researchers at major companies to accelerate the process of drug discovery. Data from some groups are used by others, and often visualizations are developed by one group for use by others. A common strategy is to start with a large pool of potential pharmaceuticals, and to have one group apply their expertise (efficacy, toxicity, drug interactions, manufacturability, etc.) to filter the list down to fewer candidates. Then they pass their results on to the next group in a sequential process.

Other companies have social structures that are more hierarchical, in which visualization tools and methods are supplied by a central office for use by all staff members. The management goal is standardized work procedures and reduced skill needs by users, who merely apply information visualizations, rather than develop new ones. This management structure is reported to be effective in a large transportation company with 225 users of treemaps and in a major military agency with 2500+ users, but a MILC evaluation would help to understand the ingredient for these reported successes and the remaining problems.

To study social creative processes MILCs could require 3-10 researchers working for 1-3 years. This level of ambition is beyond current projects, taking it closer to what is expected in clinical trials for new medications or surgical procedures. The attraction is to study complex issues such as how can individual motivation be kept high when team efforts are necessary? How can managers motivate and reward individuals and teams? What management structures are viable when large teams from diverse domains are necessary for creative endeavors?

Of course, the sobering reality is that it will often be difficult to trace successful individual or team outcomes directly to the use of a novel software tool. New results from colleagues, fresh directions for work, changes in management, and inspirations from other researchers all contribute to successful projects. In some circumstances tools are important, such as telescopes, microscopes, or computer tomography, so understanding what new tools are needed or when they should be applied may be the greatest payoffs from MILCs. We recognize that there are many limitations and concerns about MILCs, but they appear to offer an appealing way forward in studying expert users over long time periods working on complex problems. Some reassurance will come from triangulating among the multiple evaluation methods, e.g. if observations, interviews and logging data all reveal certain features to be frequently used, researchers will feel more confident in claiming value for these features.

7. CONCLUSIONS

The ambition of HCI and information visualization researchers is rising for two reasons. First they seek to study the complex patterns of work for expert users as they deal with difficult problems to produce insights, innovations, inventions, compositions, and discoveries. Assessing creativeness of work products is difficult, but no easier than assessing the creativeness of the work process. Therefore MILCs are proposed as the basic research method because its multiple methods can provide multiple perspectives on tool usage. This strategy seems to be the best hope for creating a compelling case for validity and generality, especially in situations where replicability is not attainable. The outcome may be specific suggestions for tool improvements and a better understanding of design principles. However, this paper proposes that HCI and information visualization researchers accept responsibility for a second outcome: the achievement of users' goals within their domain of work. This is a substantial increase in expectations for researchers, which raises the responsibility of researchers for the successful work of their subjects/collaborators.

A second argument of this paper is that once modest MILCs are widely applied, they should be scaled up to more ambitious projects that require 3-10 researchers working for 1-3 years, studying hundreds of users. The next step for the information visualization research community is to build the case clearly so that national funding agencies will make substantial investments in longitudinal ethnographic studies of large groups. The investment is substantial but the payoff in higher levels of creativity by more people should prove to be attractive. National funding agencies already invest heavily in clinical trials for medications, crash testing of automobiles, and space research, so we believe it is appropriate to fund major projects that study tool use in research communities. For the long run, the excitement will come from the intellectual challenge of understanding the sources of creativity and the principles for design of support tools.

Acknowledgements: We appreciate the comments from reviewers and colleagues including and Ernest Edmonds, Tom Hewett, and Jenny Preece. Partial support was provided by the U.S. National Science Foundation grant IIS-0527109 for the Workshop on Creativity Support Tools, and by the Andrew Mellon Foundation.

8. REFERENCES

[1] Amabile, T. M., *Creativity in Context*. Boulder, CO, Westview Press (1996).

[2] Argyris, C., Putnam, R., and Smith, D. M., *Action Science*. Jossey-Bass, San Francisco (1985).

[3] Bentley, R., Hughes, J., Randall, D., Rodden, T., Sawyer, P., Shapiro, D., and Sommerville, I., Ethnographically-informed systems design for air traffic control, *Proc. CSCW '92*, ACM Press, New York (1992), 123-129.

[4] Beyer, H. and Holtzblatt, K., *Contextual Design: Defining Customer-Centered Systems*, Morgan Kaufmann Publishers, San Francisco, CA (1998).

[5] Bodker, S., Scenarios in user-centered design – setting the stage for reflection and action, *Interacting with Computers 13*, 1 (2000), 61-76.

[6] Candy, L. & Edmonds, E. A., Supporting the creative user: A criteria based approach to interaction design. *Design Studies* 18 (1997), 185-194.

[7] Candy, L. and Edmonds, E.A., *Explorations in Art and Technology*, Springer-Verlag, London (2002).

[8] Dumas, J. and Redish, J., *A Practical Guide to Usability Testing: Revised Ed.*, Intellect Books, Bristol, UK (1999).

[9] Edmonds, E. A., Weakley, A. J., Candy, L., Fell, M. J., Knott, R. P. and Pauletto, S., The studio as laboratory: combining creative practice and digital technology research, *Int'l Journal of Human Computer Studies* 63, 4-5 (2005), Special Issue on Computer Support for Creativity, 452-481.

[10] Erickson, T., The design and long-term use of a personal electronic notebook: A reflective analysis, *Proc. ACM CHI 96 Human Factors in Computing Systems Conference*, ACM Press, New York (1996), 11-18.

[11] Fekete, J.-D., Grinstein, G., Plaisant, C., IEEE InfoVis 2004 Contest, the history of InfoVis, www.cs.umd.edu/hcil/iv04contest (2004)

[12] Fetterman, D. M., *Ethnography: Step by Step: 2nd Edition*, Sage, Thousand Oaks, CA (1998).

[13] Fjermestad, J. and Hiltz S.R. "Group Support Systems: A Descriptive Evaluation of Case and Field Studies," *Journal of Management Information Systems* 17, 3 (2000), 113-157.

[14] Harper, R., The organization of ethnography, *Proc. CSCW 2000*, ACM Press, New York (2000), 239-264.

[15] Hewett, T. T., The role of iterative evaluation in designing systems for usability, In M. D. Harrison and A. F. Monk (Eds.), *People and Computers: Designing for Usability* Cambridge: Cambridge University Press. (1986), 196-214.

[16] Hewett, T. T., Informing the design of computer based creativity support environments. *International Journal of Human Computer Studies* 63, 4-5 (2005), Special Issue on Computer Support for Creativity, 383-409.

[17] Hughes, J., King, V., Rodden, T., and Andersen, H., The role of ethnography in interactive systems design, *ACM interactions 2*, 2 (April 1995), 56-65.

[18] Hughes, J., O'Brien, J., Rodden, T., and Rouncefield, M., Design with ethnography: A presentation framework for design, *Proc. of Design of Interaction Systems '97*, ACM Press, New York (1997), 147–159.

[19] Lewin, K. *Field Theory in Social Science*. Basic Books, New York (1951).

[20] Marchionini, G. and Crane, G. Evaluating Hypermedia and Learning: Methods and Results from the Perseus Project, *ACM Trans. on Information Systems 12*, 1 (1994), 5-34.

[21] Millen, David, Rapid ethnography: time deepening strategies for HCI field, *Proc. ACM Symposium on Designing Interactive Systems*, ACM Press, New York (2000), 280-286.

[22] Olson, G. and Olson, J., Groupware and computer supported cooperative work, In Jacko and Sears (Editors), *The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications*, LEA Publishers, Mahwah, NJ (2002), 583-595.

[23] Plaisant, C. The Challenge of Information Visualization Evaluation, *Proc. of Conf. on Advanced Visual Interfaces*, ACM, New York (2004), 109-116.

[24] Rose, A., Plaisant, C., and Shneiderman, B., Using ethnographic methods in user interface re-engineering, *Proc. DIS '95: Symposium on Designing Interactive Systems*, ACM Press, New York (1995), 115–122.

[25] Saraiya, P., North, C., and Duca, K., An evaluation of microarray visualization tools for biological insight, *Proc. of IEEE Symposium on Information Visualization* (2004), 1-8.

[26] Seo, J. and Shneiderman, B., Knowledge discovery in high dimensional data: Case studies and a user survey for the rank-by-feature framework, *IEEE Trans. on Visualization and Computer Graphics 12*, 3 (May/June, 2006).

[27] Shneiderman, B., Creating creativity: User interfaces for supporting innovation, *ACM Transactions on Computer-Human Interaction 7*, 1 (March 2000), 114-138.

[28] Shneiderman, B., Fischer, G., Czerwinski, M., Resnick, M., Myers, B. and 13 others, Creativity Support Tools: Report from a U.S. National Science Foundation Sponsored Workshop, *International Journal of Human–Computer Interaction 20*, 2 (2006), 61–77.

[29] Shneiderman, B. and Plaisant, C., *Designing the User Interface: Strategies for Effective Human-Computer Interaction: Fourth Edition*, Addison-Wesley Publ. Co., Reading, MA (April 2004), 684 pages.

[30] Sternberg, R. (Editor), *Handbook of Creativity*, Cambridge Univ. Press, Cambridge, UK (1999).

[31] Suchman, L. A., *Plans and Situated Actions: The Problem of Human-Machine Communication*, Cambridge University Press, UK (1987).

[32] Trafton, J., Tsui, T., Miyamoto, R.; Ballas, J., Raymond, P., Turning pictures into numbers: extracting and generating information from complex visualizations. *International Journal of Human Computer Studies 53*, 5 (2000), 827-850.

[33] VAST contest: www.cs.umd.edu/hcil/VASTcontest06

[34] Warmington, A. Action Research: Its Methods and its Implications. *Journal of Applied Systems Analysis*, 7 (1980), 23-39

[35] Yin, R. K., *Case Study Research: Design and Methods, 3rd Edition*, Sage Publications, Thousands Oaks, CA. (2003).