Grounded Evaluation of Information Visualizations

Petra Isenberg*   Torre Zuk*   Christopher Collins‡   Sheelagh Carpendale*

*Department of Computer Science   Department of Computer Science‡
University of Calgary, Canada   University of Toronto, Canada
{pneumann│zuk│cmcollin│sheelagh}@cpsc.ucalgary.ca

ABSTRACT

We introduce *grounded evaluation* as a process that attempts to
ensure that the evaluation of an information visualization tool is
situated within the context of its intended use. We discuss the pro-
cess and scope of grounded evaluation in general, and then describe
how qualitative inquiry may be a beneficial approach as part of this
process. We advocate for increased attention to the field of qual-
itative inquiry early in the information visualization development
life cycle, as it tries to achieve a richer understanding by using a
more holistic approach considering the interplay between factors
that influence visualizations, their development, and their use. We
present three case studies in which we successfully used observa-
tional techniques to inform our understanding of the visual analyt-
ics process in groups, medical diagnostic reasoning, and visualiza-
tion use among computational linguists.

Categories and Subject Descriptors

H.5.2 [**Information Interfaces and Presentation**]: User Interfaces—
Evaluation/Methodology

Keywords

Information visualization, evaluation.

1. INTRODUCTION

Studying information visualization scenarios is difficult because
factors of human experience and analysis skill as well as the data
and its presentation influence performance within a given visual-
ization system. In gathering empirical evidence about the value
of information visualizations, thus far, the research community has
largely focused on performance evaluations of visualization tools
using quantitative approaches (e. g. [12], [33]). These evaluations
of information visualizations commonly occur in the controlled en-
vironments required for experiments. Since they are far removed
from actual workplaces, work flows, and often the intended users,
it is difficult to ensure transferability of results into the realities of
use in the field. As Plaisant [23] has argued, there is a need to try
and ensure that the evaluations are grounded in the context in which
they strive to assist users. This context may include such varying

factors as the size and complexity of datasets and tasks, personal
data exploration experience, current stress level, environmental dis-
tractions, cognitive processing capabilities, and analysis processes
within a given work environment. Qualitative research methodolo-
gies are designed to place an emphasis on including these reality-
based factors and thus are a natural fit for this type of inquiry. They
can be used throughout the development life-cycle. For example,
in mid- and late-stage development, qualitative techniques such as
focus groups [17], integrating heuristic assessment with experimen-
tal results [25], and expert reviews [32] have been reported. Also,
at the end of the development life-cycle, Shneiderman and Plaisant
[27] describe strategies for multi-dimensional in-depth case studies
for documenting the efficacy of visualization tools deployed in real
workplaces. Notably missing from this palette of quantitative tech-
niques as used in information visualization are examples from the
early and pre-design stages of development.

In this paper, we argue for a process that we term *grounded eval-
uation* becoming a significant part of our information visualization
evaluation practices. Grounded evaluation is a process that attempts
to ensure that the context of information visualization use grounds
further evaluations of a specific visualization tool. We use the term
evaluation broadly as the process of examining the properties of a
system or situation and also include those methods that help to as-
sess the pre-design context of information visualization without the
assessment of any digital tool. In our discussion, we focus on quali-
tative approaches including observational studies conducted as part
of the design process, in situ interviews, and field studies. We dis-
cuss the use of observational studies to create design and evaluative
criteria that are derived from data and solidly grounded in reality.
These types of studies offer potential for improved understanding
of existing practices, analysis environments, and cognitive task con-
straints as part of a grounded evaluation process.

Initially, we briefly give an overview of grounded evaluation and
how qualitative methods fit into this process. However, the purpose
of this paper is not to provide a survey of the many available qualita-
tive research methods. Rather, we discuss the context of grounded
evaluation in terms of the information visualization development
life cycle and then focus on the evaluation in the initial phase of
information visualization development, the exploration of the prob-
lem space, where *grounded evaluation* begins.

Finally, we present three case studies about our experience with
qualitative evaluations which form the first stage of grounded eval-
uation. These studies, mostly of an observational nature, are based
on grounded theory or ethnography, and ground later design and
evaluation. We hope to spark further research and application of
qualitative methods to inform our understanding of information vi-
sualization design and evaluation, and to eventually reach a wider
acceptance of this type of research methodology in our field.

2. GROUNDED EVALUATION

Grounded evaluation is a process that attempts to ensure that the evaluations of information visualizations are situated within the context of intended use. The first step in this process is to gain an understanding of the pre-design context in which information visualizations are to be used. Ideally, this understanding will include rich contextual information in terms of data, tasks, any existing visualizations, and the working methods of people for whom it is intended. This information can then be used to inform the initial design, establish context for mid-development evaluative processes such as focus groups, and help inform evaluation criteria of the resulting information visualization, by guiding choices of evaluation methods, appropriate datasets, and tasks based on the insights gained in the first step. Thus, subsequent evaluations can be grounded in the initial context that was learned.

While we applaud the fact that post-implementation evaluation is becoming more common in information visualization, we note that, as of yet, there are few empirically-based information visualization papers that focus on describing theories of visual information analysis practices. However, one approach to filling this gap in related disciplines such as Computer-Supported Cooperative Work (CSCW) and HCI, is to make use of qualitative empirical methods. This type of empirical research is often done during the pre-design stage and might be conducted in either digital or non-digital settings. For example, researchers in Computer-Supported Cooperative Work (CSCW) have learned a lot about how to support people working together *with* technology, and subsequently how to study technology use, through pre-design observation and qualitative analysis of how people work together *without* technology. The basic idea is that through observations of participants' interactions with physical artifacts a richer understanding of basic activities can be gained. This understanding can be used to hypothesize about how people will interact with technology, to inform interface design, and to guide evaluation of created designs. This approach generally relies on observation of people, inductive derivation of hypotheses via iterative data collection, analysis, and provisional verification [4].

This style of research has worked well to uncover the basic mechanics of collaborative work. For instance, Tang's study of group design activities around shared tabletop workspaces revealed the importance of gestures and the workspace itself in mediating and coordinating collaborative work [31]. Similarly, Scott et al. [26] studied traditional tabletop gameplay and collaborative design, focusing on the use of tabletop space and the sharing of items on the table, resulting in a theory about how human territoriality manifests on tables. Both studies are examples of how early evaluation can inform our understanding of work practices, which in turn can inform the design of digital systems and guide evaluation of how work practices are supported by technology. In both cases traditional, physical contexts were studied first, to understand participants' interactions with the workspace, the items in the workspace, and within the group. The results of both studies are regarded as important work in the field, providing information about how groups naturally collaborate on traditional tables and often serving as guidelines for the design and evaluation of digital systems.

Another observational study conducted to inform design and evaluation was performed by Tang and Carpendale [30]. Their study looked at the information exchange required during nurses' shift changes. The results provided a decomposition of the types of information being exchanged in various media, and potential avenues for computer support including information visualizations. In scenarios such as these, it is important to understand current practices, as new information visualizations may offer improved efficiencies

in some measured criteria, but their net impact on patient care needs to considered. For example, if a highly time efficient digital system all but eliminates temporal overlap during hand-overs then some important verbal exchanges could be lost.

While we commonly rely on the results from quantitative studies which reveal characteristics of human perception, we believe that qualitative studies of the fundamentals of information use in non-digital settings can be equally informative to information visualization research as similar studies have proved to be for CSCW and HCI. In Section 5 we report on three such studies.

Grounded evaluation begins with an understanding of a pre-design context of information visualization and uses this understanding to ground further evaluations. While in this paper we call for a greater focus on rigorous observational qualitative evaluation in information visualization as part of this process, we recognize that qualitative evaluation has been successfully used in several forms in our community. For example, Saraiya et al. used domain expert assessments of insight to evaluate bioinformatics visualizations [25], while Mazza used focus groups and semi-structured interviews in their analysis of visualization approaches to support instructors in web-based distance education [17].

3. QUALITATIVE ANALYSIS TO GROUND EVALUATION

In light of our discussion of the grounded evaluation process, we advocate increased attention to the field of qualitative inquiry, which tries to achieve a richer understanding by using a more holistic approach considering the interplay between factors that influence visualizations, their development, and their use. As qualitative analysis is still a relatively rare form of evaluation in our field, this section will introduce a very brief overview of some qualitative analysis concepts as they relate to the grounded evaluation approach.

There is a good reason why quantitative studies seem to be the most common type of evaluation in information visualization. Their strength lies in allowing the discovery of quantifiable information. Quantitative measures are often important to assess specific design choices or to compare visualization techniques *based on numbers and frequencies* like data processing and display speed, or task completion speed and accuracy. Quantitative methods provide results which can be analyzed statistically and show how results relate to hypotheses [18]. However, not all questions one may want to ask about a visualization or the context of its use can be adequately answered with quantitative approaches. When we use qualitative methods for evaluation, our goal is to collect data to *describe meaning*, rather than to make statistical inferences [4, 18]. Within the grounded evaluation process we are interested in gathering rich contextual information that may include subjective experiences.

There are a wealth of qualitative research methods that can help us to gain a rich understanding of the factors that influence information visualization use and design. These methods include action research, ethnography, conversation analysis, grounded theory (e. g. [5]), phenomenography, and many more. Qualitative methods do not need to be used exclusively, and are often complementary to quantitative methods. Gorard [8] argues that quantitative methods cannot ignore the qualitative factors of the social context of the study and the factors involved in coming to an interpretation of the results. Similarly, in qualitative research there may be factors that can be numerically recorded. These can then be presented in combination with qualitative data. For example, when one analyzes a questionnaire or interview that included both fixed-choice questions and open-ended questions, both quantitative measurement and

qualitative inquiry are typically combined [22].

3.1 Gathering Qualitative Data

Qualitative data commonly comes in one of four basic types [4, p.121]: observations, interviews, documents (written artifacts), or audio-visual materials. Acquiring these types of data can be done in a multitude of ways. In the process of gathering qualitative data there are two major concerns: determining a sampling methodology that is representative and unbiased, and when acquiring a sample the avoidance of projecting aspects or attributes onto the data that are not there (e.g. observer bias). Sample sizes for qualitative research are determined differently than for quantitative research. Since qualitative research is not concerned with making statistically significant statements about a phenomenon, the sample sizes are often lower than required for quantitative research. Often sample sizes are determined during the study by continuing qualitative inquiry until one cannot gain new data through observation [1, 18]. There is no guideline to say when this "saturation" may occur [29]. Sample sizes may vary greatly depending on the scope of the research problem but also the experience of the investigator. An experienced investigator may reach a theoretical saturation earlier than a novice investigator. Qualitative research can also be very time-consuming as it often requires large amounts of collected data to be analyzed and parsed. Therefore, compromises also have to be made between sample sizes and the amount of data that can be effectively processed.

3.2 Analyzing Qualitative Data

Qualitative data may be analyzed using qualitative, quantitative, or a combination of both methods. Mixed methods research includes a qualitative phase and a quantitative phase in the overall research study in order to triangulate results from different methods, to complement results from one method with another, or to increase the breadth and range of inquiry by using different methods [9].

Many of the qualitative analysis methods can be grouped as types of *thematic analysis*, in which analysis starts from observations, then themes are sensed through review of the data, and finally coded [2]. Coding is the process of subdividing and labeling raw data, then reintegrating collected codes to form a theory [29]. Moving from the raw data into themes and a code set may proceed using one of three approaches: data-driven, motivated from previous research, or theory-driven, each with respectively decreasing levels of sensitivity to the data [2]. In the first style, commonly called open coding [4], themes and a code set are derived directly from the data and nothing else. Theory-driven coding one may think of as a process with similar characteristics to heuristic evaluation. In either case the coded data may then be interpreted in more generalized terms. Qualitatively coded data may then then be used with quantitative or statistical measures to try and distinguish themes or sampling groups. The goal of each type of analysis process is to extract themes and to present a coherent, consistent picture of the situation under study [18].

3.3 Validity

Qualitative methods are often criticized for a lack of validity. Since the data that is analyzed in qualitative research comes from words, pictures, video, or other artifacts, qualitative researchers often do not primarily make use of statistical analysis methods. This does, however, not mean that the result of a qualitative analysis is pure speculation or based on vague impressions [18]. As explained above, qualitative data can be gathered and analyzed systematically. In qualitative research, however, it is acknowledged that the researcher's views, research context, and interpretations
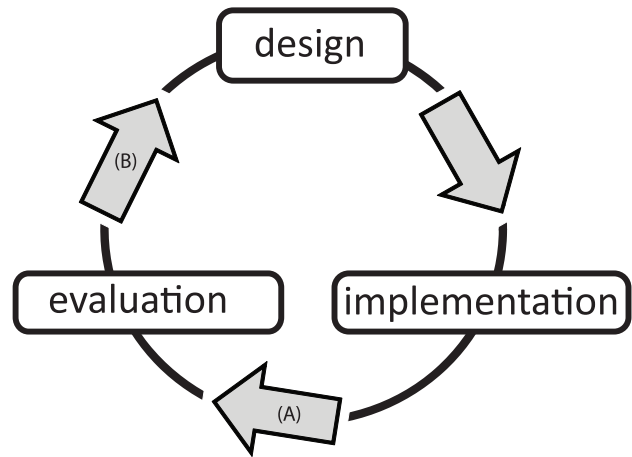

Figure 1: Evaluation in the development life cycle. Grounded evaluation advocates beginning at (A), using qualitative studies as a form of evaluation that can be carried out before initial design. Traditional infomation visualization approaches the cycle beginning at (B), designing for a specific problem after a thorough investigation of existing literature.

are an essential part of the qualitative research method as long as they are grounded in the collected data [1]. In line with our goal for grounded evaluation, the outcome of a qualitative analysis is typically very detailed, embedded in a social context, showing processes and often leading to a hypothesis or concepts about the situation under study.

4. GROUNDED EVALUATION IN THE DEVELOPMENT LIFE CYCLE

Evaluation forms a cornerstone of the full development life cycle (see Figure 1). Many information visualization researchers and designers approach the cycle beginning at Figure 1 (B), after an understanding of the design situation has been gained by reviewing related work and previous guidelines, or through informal meetings with data domain experts. We advocate a grounded evaluation approach beginning at Figure 1 (A), using qualitative studies as a form of evaluation that can be used before initial design. Initially, either problems, tasks, or the use of pre-existing visualizations need to be understood. Formal pre-design evaluative studies can be performed to gain a better understanding of the context of use and requirements for design. The main benefit of this grounding is in the rigour achieved—the exact steps leading to design decisions can be well documented and understood. Such rigour is expected at late stages, when evaluating implemented designs, but it is not common at the earliest stage. The understanding achieved through formal pre-design evaluation can be used to form criteria on which to base requirements as well as help determine user and system constraints. Once this basis for user and functional requirements and constraints has been established, design, implementation, and evaluation occur in an iterative process with the grounding being propagated into the later iterations.

While it is common to develop a visualization technique and go looking for a problem for which it may be of assistance, pre-design evaluation may assist the approach of first understanding a problem to determine what may be of assistance. Rarely do tasks and problems sit in a completely new context where no pre-existing methods or visual support exists. At this stage, qualitative meth-

ods may cover more territory than traditional quantitative methods by raising the important and often subjective issues the user faces. This is a natural fit as it is important to broaden the search space in early exploration in order to avoid committing to sub-optimal designs. Thus, this exploratory grounding, rather than being constraining, may open up the design space to consider more globally optimal solutions.

Given an early-stage grounding evaluation, we move on to the design phase. User-centered and participatory design methodologies revolve around iterative processes with the users and stake-holders. For these methods, qualitative evaluation by the participating user or domain expert is the rule rather than the exception. Quantitative aspects may be captured in new functional requirements, but user guidance is normally qualitative in nature. The investment in experimental design and numbers of participants required for many quantitative studies may be cost prohibitive at the design stage. We can rely on the outcomes of our early-stage grounding evaluation of the context in order to focus the design to the requirements and context characteristics of the problem space. The exploratory nature of a prototype is more aligned with discount evaluation methods such as heuristic evaluation, pluralistic or cognitive walk-throughs, or focus groups [19]. Given what was learned about the visualization context of use in early-stage evaluation, we can tailor these design-stage assessments to seek validation of the design. Slocum et al. [28] reported that their iterative design and evaluation process with domain experts fell short of expectations when a final evaluation with decision-makers was performed. "As it was, our evaluation tended to focus on the particular tool we had developed rather than on finding out how it might enhance their decision-making [28, p.314]." They suggest that involving the decision-makers earlier may have helped, and we interpret this as a call for a more grounded evaluation process.

Near the end of the development life cycle, before a formalized deployment, there may be some form of acceptance testing. However, tasks are often flexible and they may evolve with the introduction of new technology. So, qualitative components can be important to ensure the hypotheses are still relevant to a new workflow. Similarly, when adaptation of the user or task to the new visualization support requires long time periods, qualitative feedback may provide insight on the future acceptance or validity of the design that only long-term study will eventually provide [27]. Despite previous recommendations for in-depth long-term case studies, research implementations often do not go this pathway. However, many commercial visualization tools go into the software life cycle stage of maintenance and potential refactoring, in which continual long-term evaluations may be performed. Quantitative evaluation methods such as laboratory experiments, including usability studies, are also appropriate at this stage. Given the outcomes of the early-stage evaluations, we know in advance about issues such as perceptability of graphical variables in the intended use environment, or the charactersics of collaboration for the visualization tasks. Thus we can tailor both qualitative and quantitative post-implementation evaluation to assess these factors.

In summary, a grounded evaluation approach uses qualitative methods at pre-design stages to inform later design and evaluation. Qualitative approaches, which can be used at any stage in the development cycle, are under-used at the pre-design stage — there are only few studies that report their use to inform the design of information visualizations (e. g. [21], [24]). Figure 2 gives an example of the processes involved in performing grounded evaluation during an initial phase of development. In later phases, the diagram can be expanded in that previous evaluations can also influence the coding categories.

5. CASE STUDIES

In our research, we have successfully used grounded evaluation; specifically, we have employed various qualitative methods to inform our understanding of problem domains for which we desire new information visualizations, to guide visualization design, and to ground late-stage evaluation. In this section we will give a brief overview of how and which qualitative methods we used and how grounded evaluation benefited our research in information visualization.

5.1 An Exploratory Study of Information Analysis Processes

One of the authors' main research focus is on co-located collaboration around information visualizations. However, there is little research in this area on which to base the design and evaluation of such digital systems. We can get some advice from research on co-located groupware in the area of CSCW. In [10] we give an overview of work in CSCW which is applicable to the design of co-located collaborative information visualization systems. However, while we can learn a lot about collaborative work practices from this body of research there was little evidence about how teams engage specifically in collaborative information analysis. We do not know how and when visualizations will be shared and what types of analysis processes need to be specifically supported in collaborative information visualization systems. An observational research approach was appropriate for finding answers to these types of questions as we had no research hypothesis on which to base measurable factors for quantitative evaluations. We performed an observational study and used a grounded theory approach [29] to analyze the collected data [11]. Teams in our study were given paper-based visualizations to solve tasks, allowing us to view their process independently of the constraints of a specific information visualization system. With this setup we could observe free arrangement of data, annotation practices, and different ways of working with individual information artifacts — behaviours that might not have been implemented in similar ways in any specific information visualization system that we could have used as a work environment.

Our analysis revealed eight processes in which participants completed the tasks in our study. We showed that these eight processes relate to other models of information analysis (e. g. the sense making model [3]), and provided insights on differences and commonalities between them. We see these eight processes as a theory that can be applied to the design, heuristic evaluation, and analysis of individual and collaborative information visualization systems. A drawback of our study approach was that we would not see how typical interactions in information visualization tools (such as selection, encoding, or presentation parameter manipulations) would be used; however, our main interest was in uncovering general processes involved in collaborative and individual visual analysis rather than the low-level interactions that would be performed on information visualizations. As part of a grounded evaluation approach we will study these in further evaluations once a system for co-located collaborative information visualization has been built based on the guidelines of this initial study. We are currently working with a team of biologists at the University of Calgary for which a collaborative data analysis system will be built. As part of the grounded evaluation approach we are currently attending their weekly meetings in which experiment results are discussed to learn more about their specific collaborative data analysis practices. We will use what we learn in these weekly meetings along with our initial theory of collaborative work practices to evaluate our future software design and its use.
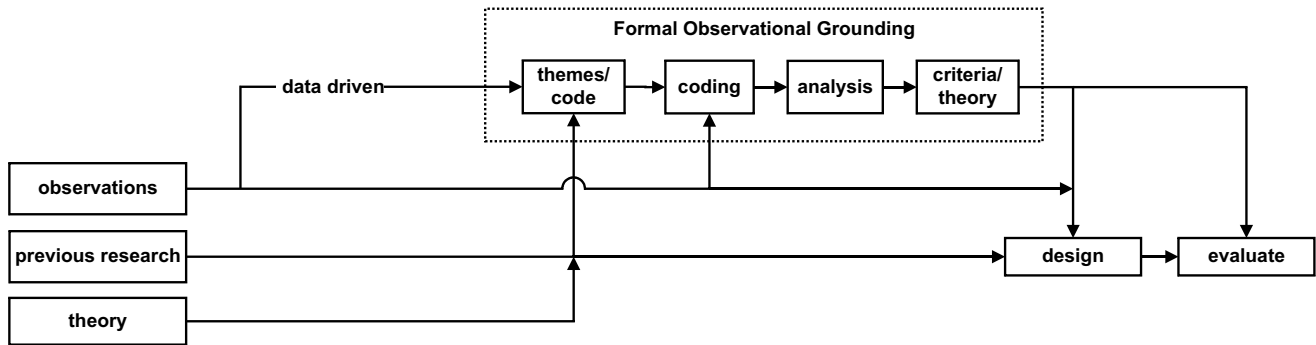
Figure 2: Grounding as a part of the first iteration of the development cycle. Currently, most information visualization development follows the path from previous research and theory to design and evaluation (bottom of diagram). Design can also be informed by observations (middle line). Observations lead directly to design, or they can be a starting point for a data-driven formal observational grounding process. If themes and codes are pre-determined, observations can lead directly to the coding stage of grounded evaluation. Otherwise, the observational data and theory can drive the development of themes and codes, starting the process of formal observational grounding. The grounding process (top of diagram) informs both design (our focus) and future evaluation.

5.2 Observational Fieldwork and Contextual Interviews in Medical Decision-making

To begin studying the challenging problem of medical diagnosis, it was important to gain an understanding of current practices around existing computer support. The use of existing custom software and the evolving practice of evidence-based medicine meant that without direct observational grounding it would be difficult to make informed design decisions. This was a natural fit for a qualitative evaluation in the form of observational fieldwork. Observing physicians in their environment working with, and working around, existing diagnostic support, allowed insights in their thought processes that would be difficult to capture with other methodologies. Cognitive processes may not be available for review on demand at a later time and so "think aloud" protocols may offer insights that vanish or are masked with retrospectives. Similarly, evaluating cognitive styles evoked only in a laboratory settings may misguide a researcher.

Observations and other qualitative results from a questionnaire and contextual interviews from the study were analyzed to discover underlying patterns and themes. Existing themes already grounded from earlier research or theory are also commonly used to analyze qualitative data [2]. The flexibility in the thematic approach allowed the application of themes from theory relating to cognition under uncertainty to be applied when looking at the data. However, this strategy may be more sensitive to bias, as one must be wary of sensing themes that are not present, which is still a concern, but lesser, if generating themes directly from the raw data.

A major benefit of the observational study was seeing the larger context of any new visualization support. The participants' motives, misgivings, and opinions shed light on how they related to existing support, effectively guiding the development of new support. For example, we determined the current decision support was provided too late in the task flow, and an existing visualization was not utilized. This type of tacit knowledge was very important at the early stage of determining what types of information visualizations may be of value, as it suggested a transparency of motives and evidence was required to allow prudent skepticism of any diagnostic recommendations. This grounding led us to a more flexible design based around visual evidence as a knowledge resource, rather than a diagnostic expert system that would only be relevant at one particular part of the task. Our resulting implementation was broadened to explore multiple visualizations spanning the larger decision context. This necessarily expanded the evaluation in order to probe what information would be relevant and how it might be presented for decision support at any point in the task [34].

5.3 Interviews and Participatory Observation in Computational Linguistics

We have also used observational fieldwork and contextual interviews to better understand the domain-specific problems faced by computational linguists developing machine translation systems. The problem space we were investigating was complex: the linguists' data consists of hundreds of thousands of sentence pairs (same sentence in two languages) and millions of learned translation rules; the translation algorithms run on hundreds of processors in a distributed computing grid; and the team is made up of more than 10 researchers in several locations. Preliminary interviews revealed that they spent most of their time sitting at a computer, programming. The data analysis parts of their work — examining training data and testing output — come in unpredictable phases with two weeks or more of coding between cycles. This did not lend itself to long term observational study, so a series of contextual interviews was conducted in their research environment in order to better understand the work situation.

Our interview protocol was designed to investigate many areas of the data, tasks, and work practices: we tried to be as broad as possible, investigating beyond our assumptions of where we thought visualization is generally helpful. We investigated the individual researcher's understanding of their broader research project, their understanding of the algorithms and data, their analysis tools (including ad hoc visualizations, white board sketches, and notebooks) and practices (using a cognitive walk-through of a typical analysis), their collaboration practices and collaboration support tools, and the ways they measured the success of their research.

We followed up these interviews with two days of participatory observation [14]. In participatory observation, additional insights can be gained through first-hand experience of the data analysis process and context of study. Our observer had some training in computational linguistics, so was well-suited to this task, but domain expertise is not essential for successful participatory observation. The study participants trained our observer on typical data analysis tasks for several hours, and then "put him to work" on data analy-

sis using their existing tools and techniques. The observer kept a journal of the experience, and the outcomes of the analysis were reviewed with the domain experts for validity afterward. Most of the observations were considered valid and significant by the domain experts. We are currently reviewing collected artifacts and using open coding [2] on transcripts, field notes, and the participant observation journal to generate a rich understanding of the data.

Early results reveal a surprisingly sophisticated reliance on visualizations (custom-made digital visualizations and hand drawings) in a research group unfamiliar with information visualization research. We see opportunities to build interactive visualizations to support not only the analysis of their translation data, but also to support understanding of how their distributed algorithms work, and to support collaborative analysis of intermediate states in the translation process. If we had relied on domain expert requests, simple task and requirements analysis techniques, or our own assumptions, we likely would be constrained for designing visualizations for single users on domain data, where that may not be the area of most potential success. Now that we are moving forward with a better grounding in the proposed context of visualization use and the problem space, our designs and post-implementation evaluation will focus on algorithm and data visualization for collaborative analytics conducted in small groups.

5.4 Summary of Case Studies

While this paper frames the process of integrating qualitative empirical research from the pre-design phase onward as grounded evaluation, the technique has already been used in several of our recent and ongoing research projects. Similarly to the results obtained from observational studies employed in other domains, we are finding this approach, while time and resource intensive, to be informative and beneficial. The first case study, which considered the visual analytics as conducted by small groups using pencil-and-paper, illustrates how the richer understanding developed through observational study can provide insights that go beyond the constraints of existing digital technology to reveal how participants could make use of additional freedoms. Specifically, we observed that when people were not constrained by temporal sequencing already present in information software, they were extremely varied and inventive in their choice of process sequencing.

Our second case study, observing medical diagnostic reasoning in situ followed by contextual interviews, revealed a complex interplay between uncertainty in data and uncertainty in diagnostic processes. This led not only to insight into the development of appropriate software support tools but also prompted a further investigation into uncertainty in reasoning and a broader consideration of the impact of uncertainty visualization in general.

Our third case study, field interviews and observations in computational linguistics, has provided an array of new understanding into the practical invention and use of visualization being made by people to help address their own complex intellectual tasks. All these studies bridged the divide between ourselves as visualization experts and the domain experts for whom we are creating visualizations. That is, while we have not become analysts, doctors, or linguists, we have gained a deeper understanding of how they work and think. This can open up the design space revealing new possibilities for information visualizations, as well as additional criteria on which to measure success.

6. MOVING FORWARD

While there is growing recognition in our research community that it is difficult to evaluate information visualizations (e. g. [20, 23, 27]), the recognition of this difficulty has not in itself provided immediate answers of how to approach this problem.

Recent positive trends of note include: one, that more quantitative evaluative papers in the form of usability studies are being published (e. g. [12, 16, 33]), two, that there are several papers that have made a call for more qualitative evaluations and complementary qualitative and quantitative approaches (e. g. [6, 17, 32]), and three, a few, albeit still too few, qualitative evaluations of information visualizations have been published (e. g. [7, 15, 25]). While the tide appears to be turning, we, as a research community, are still not sufficiently aware of the variety of qualitative studies possible and the potential benefits they can offer to the research community. Also, there maybe uncertainty about how to judge the quality of a qualitative evaluation paper, leading to difficulty in reviewing. In this paper we have focused on a small part of this problem, primarily discussing the use of qualitative research methods as part of the design process and the way this can feed back into the ongoing evaluative process. The techniques discussed in the case studies have been developed through borrowing from ethnographic and sociological research methods, and applying them to our information visualization needs. Pre-design qualitative studies are capable of providing us with rich description of the situation of interest or, in other words, an in depth understanding of the context, task, and processes involved with the type of information tasks we are looking to support through visualization.

The types of observational studies that are done to inform design typically focus on a particular situation, such as in our case studies on visual information analysis, medical diagnosis, and computational linguistic analysis. However, they can provide useful generalizable insights. Examples of this type of general contribution can be found in the CSCW literature. For instance, studying pairs of people assembling puzzles [13], led to an understanding of how variant orientation is important during collaboration for communication and coordination purposes. Observing people during graphic design sessions [31] led to series of insights to the collaborative process that has informed much of the subsequent CSCW research. Thus, not only is it important to conduct initial observational studies as part of one's design process, it is also important to publish these studies because they will inform and enrich the research community in general.

7. CONCLUSIONS

We introduce *grounded evaluation* as a process that attempts to ensure that the evaluations of information visualizations are situated within the context of intended use. Reports of formal evaluations of information visualization contexts can not only be extremely beneficial during the design phase but also offer valuable information on how to evaluate and measure the successes or failures of information visualization tools. Qualitative studies can be a powerful method by which one can capture salient aspects of a problem; they can provide useful design and evaluation criteria. Quantitative evaluation is naturally precision-oriented, but a shift from high precision to high fidelity may be made with the addition of complementary qualitative evaluations. Qualitative evaluations can be used throughout the entire development life cycle. In other research areas, such as CSCW, and in our own case studies, we have found that observational studies to inform design are especially useful, yet these techniques are under-used and under-reported in the information visualization literature. A broader scope and sensitivity to context are important when complex issues such as collaboration, insight, and confidence, need to be assessed. In more general terms, we would like to draw attention to qualitative research approaches early in the development to address difficult evaluation questions later in the process. Just as the sign in Albert Einstein's

office read, "Everything that can be counted does not necessarily count; everything that counts cannot necessarily be counted", and thus for evaluation to be effective when we determine the things that count, we should not leave out those things that cannot be counted.

8. ACKNOWLEDGMENTS

9. REFERENCES

[1] C. F. Auerbach. *Qualitative Data: An Introduction to Coding and Analysis*. New York University Press, New York, 2003.

[2] R. E. Boyatzis. *Transforming Qualitative Information: Thematic Analysis and Code Development*. Sage Publications, London, 1998.

[3] S. Card, J. D. Mackinlay, and B. Shneiderman, editors. *Readings In Information Visualization: Using Vision To Think*. Morgan Kauffman Publishers, San Francisco, USA, 1999.

[4] J. W. Creswell. *Qualitative Inquiry and Research Design: Choosing Among Five Traditions*. Sage Publications, London, 1998.

[5] I. Dey. *Grounding Grounded Theory: Guidelines for Qualitative Inquiry*. Academic Press, San Diego, London, 1999.

[6] G. Ellis and A. Dix. An explorative analysis of user evaluation studies in information visualization. In *Proc. of the workshop on BEyond time and errors: NovelEvaluation Methods for Information Visualization (BELIV)*, 2006.

[7] V. González and A. Kobsa. A workplace study of the adoption of information visualization systems. In *Proceedings of the International Conference on Knowledge Management*, pages 92–102, 2003.

[8] S. Gorard. *Combining Methods in Educational Research*. McGraw-Hill Education, New York, USA, 2004.

[9] J. C. Greene, V. J. Caracelli, and W. F. Graham. Toward a conceptual framework for mixed-method evaluation design. *Educational Evaluation and Policy Analysis*, 11(3):255–274, 1989.

[10] P. Isenberg and S. Carpendale. Interactive tree comparison for co-located collaborative information visualization. *IEEE Transactions on Visualization and Computer Graphics (Proceedings of IEEE Conference on Information Visualization)*, 12(5), Nov/Dec 2007.

[11] P. Isenberg, A. Tang, and S. Carpendale. An exploratory study of visual information analysis. In *Proceedings of the Conference on Human Factors in Computing Systems (CHI)*, New York, NY, USA, 2008. ACM Press. To appear.

[12] A. Kobsa. User experiments with tree visualization systems. In *Proceedings of the IEEE Symposium on Information Visualization*, pages 9–26, 2004.

[13] R. Kruger, S. Carpendale, S. D. Scott, and A. Tang. Fluid integration of rotation and translation. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*, pages 601–610, New York, USA, 2005. ACM Press.

[14] C. Lewin, editor. *Research Methods in the Social Sciences*. Sage Publications, London, 2004.

[15] G. Mark, K. Carpenter, and A. Kobsa. A model of synchronous collaborative information visualization. In *Proceedings of the Conference on Information Visualization (IV)*, pages 373–381, Los Alamitos, USA, 2003. IEEE Press.

[16] G. Mark, A. Kobsa, and V. Gonzalez. Do four eyes see better than two? Collaborative versus individual discovery in data visualization systems. In *Proceedings of the Conference on Information Visualization (IV)*, pages 249–255, Los Alamitos, USA, July 2002. IEEE Press.

[17] R. Mazza and A. Berrè. Focus group methodology for evaluating information visualization techniques and tools. In *Proceedings of the Conference on Information Visualization (IV)*, 2007.

[18] W. L. Neuman. *Social Research Methods: Qualitative and Quantitative Approaches*. Allyn & Bacon, third edition, 1997.

[19] J. Nielsen and R. L. Mack. *Usability Inspection Methods*. John Wiley & Sons, 1994.

[20] C. North. Toward measuring visualization insight. *IEEE Computer Graphics and Applications*, 26(3):6–9, 2006.

[21] E. S. Patterson, D. Tinapple, E. M. Roth, J. M. Finley, and G. G. Kuperman. Aiding the intelligence analyst in situations of data overload: From problem definition to design concept exploration. Technical Report ERGO-CSEL 01-TR-01, Institute for Ergonomics/Cognitive Systems Engineering Laboratory Report, 2001.

[22] M. Q. Patton. *Qualitative Research and Evaluation Methods*. Sage Publications, London, third edition, 2001.

[23] C. Plaisant. The challenge of information visualization evaluation. In M. F. Costabile, editor, *Proceedings of the Working Conference on Advanced Visual Interfaces (AVI)*, pages 109–116, New York, USA, 2004. ACM Press.

[24] D. M. Russell, M. J. Stefik, P. Pirolli, and S. K. Card. The cost structure of sensemaking. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*, pages 269–276, New York, USA, 1993. ACM Press.

[25] P. Saraiya, C. North, and K. Duca. An insight-based methodology for evaluating bioinformatics visualizations. *IEEE Transactions on Visualization and Computer Graphics (Proceedings of the IEEE Symposium on Information Visualization)*, 11(4):443–456, 2005.

[26] S. D. Scott, M. S. T. Carpendale, and K. M. Inkpen. Territoriality in collaborative tabletop workspaces. In *Proceedings of the ACM Conference on Computer-Supported Cooperative Work (CSCW)*, pages 294–303, New York, USA, 2004. ACM Press.

[27] B. Shneiderman and C. Plaisant. Strategies for evaluating information visualization tools: Multi-dimensional in-depth long-term case studies. In *Proc. of the workshop on BEyond time and errors: NovelEvaluation Methods for Information Visualization (BELIV)*, 2006.

[28] T. A. Slocum, D. C. Cliburn, J. J. Feddema, and J. R. Miller. Evaluating the usability of a tool for visualizing the uncertainty of the future global water balance. *Cartography and Geographic Information Science*, 30(4):299–317, 2003.

[29] A. Strauss and J. Corbin. *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*. Sage Publications, London, 1998.

[30] C. Tang and S. Carpendale. An observational study on information flow during nurses' shift change. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*, pages 219–228, New York, USA, 2007. ACM.

[31] J. C. Tang. Findings from observational studies of collaborative work. *International Journal of Man-Machine Studies*, 34(2):143–160, February 1991.

[32] M. Tory and T. Möller. Evaluating visualizations: Do expert reviews work? *IEEE Computer Graphics and Applications*, 25(5):8–11, Sept./Oct. 2005.

[33] W. Willett, J. Heer, and M. Agrawala. Scented widgets: Improving navigation cues with embedded visualizations. *IEEE Transactions on Visualization and Computer Graphics (Proceedings of IEEE Conference on Information Visualization)*, 12(5), Nov/Dec 2007.

[34] T. D. Zuk. *Visualizing Uncertainty*. PhD thesis, University of Calgary, Calgary, Alberta, 2008.