Communication and Organization in Software Development: An Empirical Study

Carolyn B. Seaman
Victor R. Basili

University of Maryland
Institute for Advanced Computer Studies
Computer Science Department
College Park, MD, USA

Abstract

The empirical study described in this paper addresses the issue of communication among members of a software development organization. The independent variables are various attributes of organizational structure. The dependent variable is the effort spent on sharing information which is required by the software development process in use. The research questions upon which the study is based ask whether or not these attributes of organizational structure have an effect on the amount of communication effort expended. In addition, there are a number of blocking variables which have been identified. These are used to account for factors other than organizational structure which may have an effect on communication effort. The study uses both quantitative and qualitative methods for data collection and analysis. These methods include participant observation, structured interviews, and graphical data presentation. The results of this study indicate that several attributes of organizational structure do affect communication effort, but not in a simple, straightforward way. In particular, the distances between communicators in the reporting structure of the organization, as well as in the physical layout of offices, affects how quickly they can share needed information, especially during meetings. These results provide a better understanding of how organizational structure helps or hinders communication in software development.

1. Introduction

Software development managers strive to control all of the factors that might impact the success of their projects. However, the state of the art is such that not all of these factors have been identified, much less understood well enough to be controlled, predicted, or manipulated. One factor that has been identified [Curtis88] but is still not well understood is information flow. It is clear that information flow impacts productivity (because developers spend time communicating) as well as quality (because developers need information from each other in order to carry out their tasks well). The study described in this paper addresses the productivity aspects of communication by empirically studying the organizational and process characteristics which influence the amount of effort software developers spend in communication activities. This is a first step towards providing management support for control of communication effort.

This research also arises out of an interest in the organizational structure of software enterprises and how it affects the way software is developed. Development processes affect, and are affected by, the organizational structure in which they are executed.

This work was supported in part IBM's Centre for Advanced Studies, and by NASA grant NSG-5123.

Communication in software development is one area in which organizational and process issues are intertwined. A process requires that certain types of information be shared between developers and other process participants, thus making information processing demands on the development organization. The organizational structure, then, can either facilitate or hinder the efficient flow of that information.

The empirical study described here aims to identify the organizational characteristics which affect process communication effort, and to determine the degree of effect. The dependent variable in this study is communication effort, defined as the total effort expended to share some type of information. The dependent variables are organizational distance, physical distance, and familiarity. All three of these are measures of the organizational structure, defined as the network of relationships between members of the software development organization. The types of relationships upon which these measures are based are, respectively, official relationships, physical proximity, and past and present working relationships.

The study combines quantitative and qualitative research methods. Qualitative methods are designed to make sense of data represented as words and pictures, not numbers [Gilgun92]. Qualitative methods are especially useful when no well-grounded theories or hypotheses have previously been put forth in an area of study. Quantitative methods are generally targeted towards numerical results, and are often used to confirm or test previously formulated hypotheses. They can be used in exploratory studies, but only where well-defined quantitative variables are being studied. We combine these paradigms in order to flexibly explore an area with little previous work, as well as to provide quantified insight that can help support the management of software development projects.

The purpose of this report is twofold. First, we wish to describe the methods we have used to carry out this study, so that other researchers can consider their appropriateness for investigating this area. Second, we wish to present a set of useful results, in order for practitioners and others to gain more understanding of communication and organizational issues in software development projects. In the subsections which follow, the specific problem addressed by this study is presented, as well as the research questions and some definitions of terms. In section 2, the related work in the literature is outlined. Our research methods are described in detail in section 3, and section 4 presents our results. In section 5, some of the limitations of this study are presented and packaged as experience to be used in future efforts to address this issue. Finally, section 6 discusses and summarizes the results of the study.

1.1. Problem Statement

Software development organizations do not currently know how to ensure an efficient flow of information among developers. They do not know how to assess, with any certainty, the information flow requirements of the development processes they choose. In addition, they do not have a deep understanding of how their organizational context affects the level of effort needed to meet the process's communication requirements.

The lack of understanding of communication issues has several consequences. First of all, managers have no way to account for communication costs in their project planning, or to balance those costs with the benefits of communication. Additionally, they do not know how to identify or solve communication problems when they arise. Finally, we cannot begin to learn from experience about communication issues until we identify the important variables that affect communication efficiency.

1.2. Research Questions

The study of organizational issues and communication in software development is not advanced to the point where it is possible to formulate well-grounded hypotheses. Therefore, this work is based on the following set of research questions:

- How does the distance between people in the management hierarchy of a software development organization affect the amount of effort they expend to share information?
- How does a group of software developers' familiarity with each other's work affect the amount of effort they expend to share information?
- How does the physical distance between people in a software development organization affect the amount of effort they expend to share information?

These questions are all operationally specialized versions of the more general question:

- How does the organizational structure in which software developers work affect the amount of effort it takes them to share needed information?

These research questions lead directly to a set of dependent and independent variables for the study proposed in this document. The dependent variable is *Communication Effort*, defined as the amount of effort expended to complete an interaction. Secondly, there is a set of independent variables which represent organizational structure. Three different measures have been chosen which capture the different properties mentioned in the first three research questions above. The first, *Organizational Distance*, measures the distance between people in the official management structure of the development organization. The second is *Familiarity*, which reflects how familiar different developers are with each others' past and present work. Finally, the independent variable *Physical Distance* is a measure of physical proximity. The proposed study will explore the relationship between each of these three independent variables, and the dependent variable.

The study design also includes a large set of intervening, or blocking, variables. These factors are believed to have an effect on communication effort, but are not the primary concern of this study.

1.3. Definitions

In this section, some important concepts are defined in the context of this study.

organizational structure - the network of all relevant relationships between members of an organization. These relationships may affect the way people perform the process at hand, but they are not defined by the process being performed.

process - a pre-defined set of steps carried out in order to produce a software product.

process communication - the communication, between members of a development project, required explicitly by a software development process.

communication effort - the amount of effort, in person-minutes, expended to complete an interaction, including the effort spent requesting the information to be shared, preparing the information, transmitting or transferring the information from one party to another, and digesting or understanding the information. This definition includes effort spent on activities not normally considered "communication" activities (e.g.preparing and reading written information).

interaction - an instance of communication, in which two or more people are explicitly required (by the process they are executing) to share some piece of information. For example, the handoff of a coded component from a developer to a tester is an interaction. One developer asking for advice from an expert, no matter how crucial that advice may be, is not an interaction according to our definition. An interaction begins when some party requests information (or when a party begins preparation of unrequested information) and ends after the information has been received and understood sufficiently for it to be used (e.g. read).

qualitative data - data represented as words and pictures, not numbers [Gilgun92]. **Qualitative analysis** consists of methods designed to make sense of qualitative data.

quantitative data - data represented as numbers or discrete categories which can be directly mapped onto a numeric scale. **Quantitative analysis** consists of methods designed to summarize quantitative data.

participant observation - research that involves social interaction between the researcher and informants in the milieu of the latter, during which data are systematically and unobtrusively collected [Taylor84].

structured interviewing - a focused conversation whose purpose is to elicit responses from the interviewee to questions put by the interviewer [Lincoln85].

coding - a systematic way of developing and refining interpretations of a set of data [Taylor84]. In this work, coding refers specifically to the process of extracting specific pieces of information from qualitative data in order to provide values for quantitative research variables.

triangulation - the validation of a data item with the use of a second data source, a second data collection mechanism, or a second researcher [Lincoln85].

member checking - the practice of presenting analysis results to members of the studied organization in order to verify the researcher's conclusions against the subjects' reality [Lincoln85].

2. Related Work

The work proposed in this document is supported by the literature in three basic ways. First of all, the research questions in section 1.2 have been raised in various forms in the literature. The relationship between communication and organizational structure (in organizations in general) is a strong theme running through the organization theory literature, from classic organization theory [Galbraith77, March58], to organizational growth [Stinchcombe90], to the study of technological organizations [Allen85], to business process reengineering [Davenport90, Hammer90]. This relationship has not been explored in detail in software development organizations. However, several studies have provided evidence of the relevance of both organizational structure (along with other "non-technical" factors) and communication in software development. In particular, at least one study [Curtis88, Krasner87] points to the three aspects of organizational structure which we address in our research questions.

Second, our chosen dependent and independent variables have all appeared in some form in the literature. Our dependent variable, Communication Effort, has been defined to include both technical and managerial communication, to reflect only that communication required by the development process, but to include *all* such process communication. These decisions are based on results presented in the organization theory [Ebadi84, MaloneSmith88] and empirical software engineering literature [Ballman94, Bradac94, Perry94]. The three independent variables also appear in these two areas of literature. Organization theory points to the benefits of organizational and physical proximity of communicators [Allen85, Mintzberg79], while empirical software engineering has shown the drawbacks of organizational and physical distance [Curtis88]. The idea of "familiarity" is referred to in a more general way in the literature. Both areas refer to the importance of various types of relationships between communicators [Allen85, Curtis88]. In particular, a software development study [Krasner87] has discovered the importance of "shared internal representations", which have led to our particular definition of Familiarity.

Third, the literature in many areas has helped shape the design of the proposed study by providing methods and experience. The choice and definition of the unit of analysis, the interaction (section 3.2), has been influenced by the organization theory [Ebadi84, Liker86] and empirical software engineering literature [Perry94]. The scope of the study, in terms of the types of communication studied, has also been influenced by this literature. Data collection and analysis methods have come directly from the literature in empirical methods [Lincoln85, Taylor84].

Despite the considerable support in the literature, there are several significant issues which are not addressed there. Probably the most important is that of intervening, or blocking, variables. Our own experience and intuition strongly suggest that the influence of organizational structure on communication effort is neither direct nor exclusive. There are other factors that affect the amount of time and effort an interaction takes. We have relied on our own experience, and on conversations with many experienced managers, developers, and researchers at the study site, to identify these factors.

Another issue which is not resolved in the literature is a satisfactory way of modeling a process in terms of its individual interactions. In many cases, a process model or definition document will be written in such a way that the required interactions (as defined in section 1.3) are clearly defined. But even when this is the case, it is not clear that the model accurately reflects reality. What rules exist for separating one interaction from another? The breakdown of interactions presented in section 3.3.3 went through a number of iterations until we found a model that both reflected reality and which facilitated the collection of data.

From research questions to variables to research design, the proposed work is supported in the literature. We have extended the current state of the literature not only by combining pieces that have not previously been combined, but also by adding new approaches that were necessary to adequately address the issues of interest.

3. Research Methods

This empirical study examines the role of organizational structure in process communication among software developers. This section explains in detail the methods we employed to investigate this issue. In the subsections which follow, we first present an overview of the research plan and a discussion of our unit of analysis. Then we describe the setting of the study. Finally, the details of data collection, coding, and analysis are presented.

3.1. Overview

Our research design combines qualitative and quantitative methods. There are a number of ways in which such methods have been combined in studies in the literature. The practice adopted for this study is to use qualitative methods to collect data which is then quantified, or coded, into variables which are then analyzed using quantitative methods. Examples of this practice are found in [Sandelowski92, Schilit82, Schneider85].

The data collection procedures used in this study are participant observation [Taylor84] and structured interviews [Lincoln85]. Development documents from the environments under study also provide some of the data [Lincoln85]. As described later, the data gathered from these different sources overlaps, thus providing a way of triangulating [Lincoln85], or cross-checking the accuracy of the data.

After the data was collected, it was coded in such a way that the result is a set of data points, each of which has a set of values corresponding to a set of quantitative research variables. For example, although participant observation in general yields qualitative (non-quantified) data, we used this data to count the number of people present at the observed meeting, to time the different types of interactions that take place, and to determine what type of communication medium was used. These quantified pieces of data constitute values for the research variables.

The data analysis part of the research design is mostly quantitative. The coded data set was analyzed using very simple statistical methods. Histograms were constructed to determine the distributions of single variables, and scatterplots were used to study the relationships between pairs of variables. Various subsets of the data were also viewed in this way in order to gain a deeper understanding of the findings.

3.2. Unit of Analysis

A brief discussion of the unit of analysis is in order. The unit of analysis in this study is the *interaction*. In section 1.3, an interaction was defined as an instance of communication, in which two or more people are explicitly required (by the process they are executing) to share some piece of information. It should be noted that only process-oriented interactions are considered in this study. For example, a document handoff between different sets of developers, a review meeting, and a joint decision on how to proceed are all considered interactions in this context if they are required as part of some defined process step. We would not include, for example, informal (optional) consultations on technical matters between developers, even though this type of communication might be "required" because a developer cannot accomplish a given task adequately without it. Such informal communication is a very important, but we believe separate, area of research.

Most social science research methods assume that the unit of analysis is a person, and methods are described with this assumption, at least implicitly. However, there are some examples in the literature of empirical studies which use a unit of analysis other than a person or group of people. One is Schilit's [Schilit82] study of how workers influence the decisions of their superiors in organizations. The unit of analysis in this study is called an interaction, in this case an attempt by a subordinate to influence a superior in some way. The research variables represented characteristics of such interactions, e.g. method of influence. Like our study, this is an example in which the unit of analysis is a phenomenon, or an event, rather than a person. Other examples of non-human units of analysis can be found in the research literature on group therapy.

There are several ramifications of using this type of unit of analysis. First of all, the "size" of the study cannot be stated in terms of people. The number of people interviewed or observed is not a meaningful size measure since not all people involved in interactions are interviewed, and the same person may be present in a number of observations. The size of the study is *the number of interactions*. Each interaction constitutes one data point, and all of the variables are evaluated with respect to an interaction.

Another possible complication with this unit of analysis is the problem of independence. Any analysis method that attempts to describe a relationship between variables has an underlying requirement that the values (of variables) associated with one data point are not in any way dependent on the values associated with another data point. It can be argued that, since different interactions can involve the same people, they may not be independent. It's not clear, however, that independence has this meaning when the unit of analysis is not a person. It can be argued that the properties of people that are relevant in our context are represented as variables, and thus any dependence between two data points simply means that they share the same values for some variables. In any case, this issue should be taken into account when assessing the results reported in section 4.

3.3. Study Setting

This study took place at IBM Software Solutions Laboratory in Toronto, Canada. The development project studied was DB2, a commercial database system with several versions for different platforms. During the month of June 1994, data was collected from (mostly design and code) reviews in Toronto. Ten reviews were directly observed, which involved about 100 interactions. These observations were followed up with interviews with review participants in November 1994, and in April 1995. The review process was chosen for study because it is well-defined in the DB2 project, it involves a lot of communication between participants, and much of it is observable.

A three-part model of the DB2 development environment was built. The model had several purposes. First, it was used to better understand the DB2 review process and the people involved. Also, it served as a vehicle with which to communicate with developers and others from whom we were collecting information. Finally, it was used as a framework in which to organize the data.

Recall that the issue which motivates this work is the relationship between development organizations and processes. Information flow is one area in which organizational and process issues come together. To reflect this, the model of the DB2 environment is organized in three parts. One part corresponds to the development process under study, one to the organizational structure in which that process is executed, and one to the intersection between the two, which is modeled as a set of interactions. In section 1.3, we defined an interaction, as it is used here, as an instance of communication in which two or more process participants must share some piece of information in order to carry out their process responsibilities. The three parts of the model are described in sections 3.3.1 through 3.3.3.

3.3.1. Process

The work that goes into each release of DB2 is divided into *line items*, each of which corresponds to a single enhancement, or piece of functionality. Work on a line item may involve modification of any number of software components. For each line item, reviews are conducted of each major artifact (requirements, design, code, and test cases). In this study, we observed and measured reviews of all types, but mostly design and code.

The review process consists of the following general steps:

Planning - The Author and Line Item Owner (often the same person) decide who should be asked to review the material. The Author then schedules the review meeting and distributes the material to be reviewed.

Preparation - All Reviewers read and review the material. Some Reviewers write comments on the review material, which they later give to the Author. The Chief Reviewer sometimes checks with each Reviewer before the meeting to make sure they have reviewed the material.

Review Meeting - There are a number of different ways to work through the material during the meeting, and to record the defects. In some cases, the Moderator records all defects raised on Major Defect Forms, which were given to the Author at the end of the meeting. In other reviews, the Moderator does not use the forms, but writes down detailed notes of all defects, questions, and comments made. In still others, the Moderator takes only limited notes and each Reviewer is expected to provide written comments to the Author. In all cases, the Reviewers make a consensus decision about whether a re-review is required. Also, at the end of the meeting, the Moderator fills out most of the Review Summary Form and gives it to the Chief Reviewer.

Rework - The Author performs all the required rework.

Follow up - The Chief Reviewer is responsible for making sure that the rework is reviewed in some manner. This could take place in an informal meeting between the Author, Chief Reviewer, and sometimes the Line Item Owner. In other cases, the Author simply gives the Chief Reviewer the reworked material and the Chief Reviewer reviews it at his or her convenience. After the rework is reviewed, the Chief Reviewer completes the Review Summary Form and submits it to the Release Team.

3.3.2. Organization

The formal DB2 organization has a basic hierarchical structure. First-line managers are of three types. Technical managers manage small teams of programmers who are responsible for maintaining specific collections of software components. Groups of developers reporting to a technical manager may be further divided by component and Task Leaders may be assigned to head each subgroup. Product managers are responsible for managing releases of DB2 products. Teams reporting to product managers coordinate all the activities required to get a release out the door. Support managers manage teams that provide support services, like system test, to all the other teams. There is one second-line manager responsible for all DB2 development.

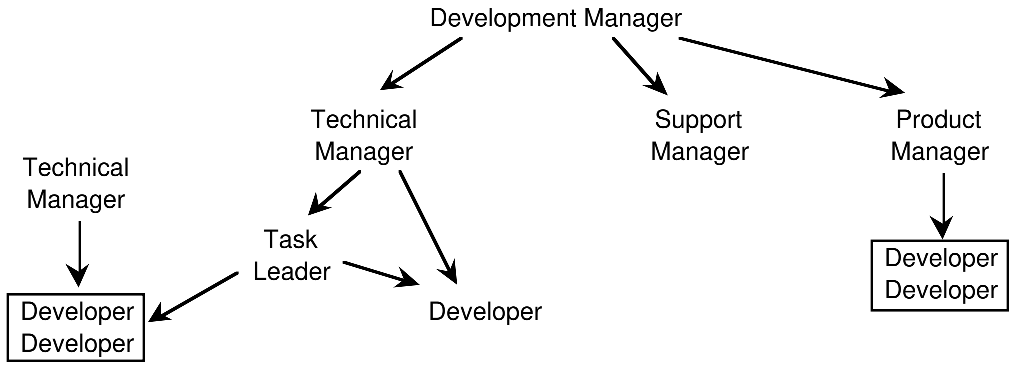

Figure 1. Example reporting relationships.

The part of the three-part model which depicts the organizational structure of the DB2 team consists of a set of simple graphs. Each graph shows a different perspective on the organizational structure. In each, the nodes are the people that constitute the team or are relevant to the development process in some way. The edges or groupings show different relationships between the members of the organization. One graph (an example appears in Figure 1) shows the reporting relationships, and is derived from the official organization chart. Note that the reporting structure need not be strictly hierarchical, and official relationships other than the traditional manager/employee relationship can be represented (e.g., the "Task Leader"). Another (Figure 2) shows the same organization members linked together according to work patterns. Those people that work together on a regular basis and are familiar with each others' work are linked or grouped. A third graph reflects the physical locations of the members of the organization (Figure 3). People who share offices, corridors, buildings, and sites are grouped with different types of boxes. These graphs are used to measure several properties of the relationships between organizational members (see section 3.5).

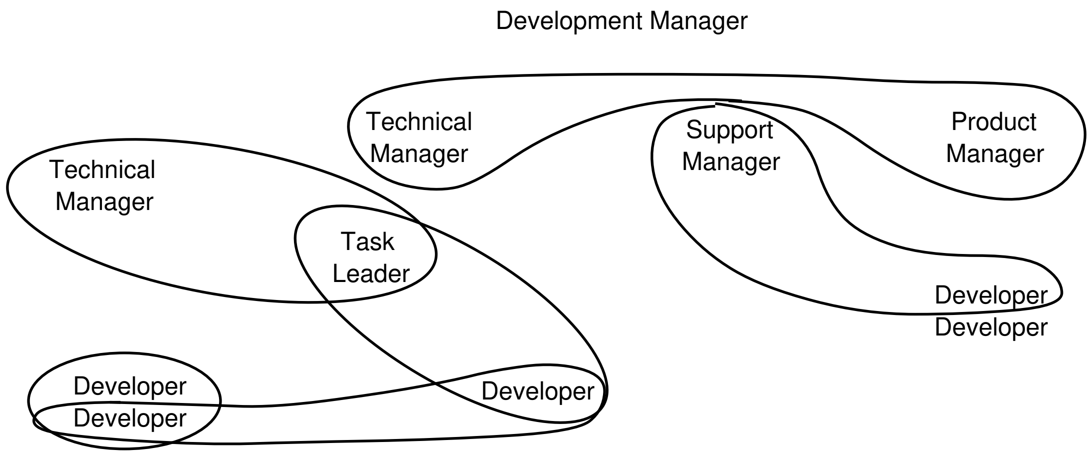

Figure 2. Example working relationships.

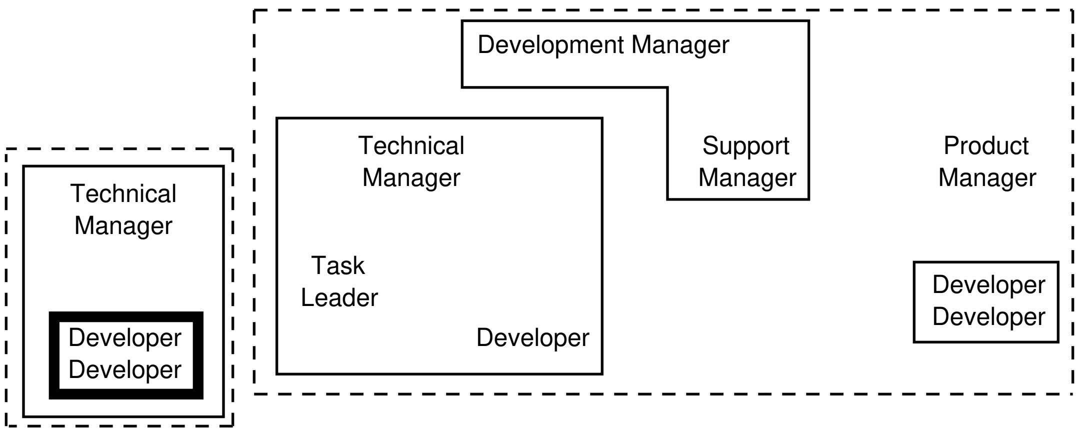

Figure 3. Example physical proximity relationships

3.3.3. Interactions

The third part of the model of the DB2 environment is made up of the types of interactions, or instances of communication, that are both dictated by the defined software development process and that actually take place, between members of the organization. These interactions constitute the overlap, or relationship, between the DB2 process and organization.

Each interaction has a set of participants. Interactions also have a *mode* of interaction, which restricts which participants interact with which others. If an interaction is *multidirectional*, then the set of participants is considered one large set, and all participants interact with all other participants. Information flows in all directions between all participants. If an interaction is *unidirectional* or *bidirectional*, then the participants are divided into two sets. Participants in one set interact only with those in the other set. In the case of unidirectional interactions, information flows from one set to the other. In bidirectional interactions, information flows in both directions.

Below are the types of interactions we have identified as *potentially* occuring during any review. Type names are meant to describe the information that is shared during the interaction:

choose_participants - the Author and Line Item Owner choose the Reviewers
review_material - the Author gives the material to be reviewed to the Reviewers
preparation_done - the Chief Reviewer asks each Reviewer if they have completed reviewing the material
schedule_meeting - the Author schedules the review meeting at a time convenient to all Reviewers
commented_material - one or more Reviewers give copies of the reviewed material, with their written comments on it, to the Author
comments - the Moderator gives the comments he or she has recorded during the review meeting to the Author
summary_form - the Moderator gives the partially completed Review Summary Form to the Chief Reviewer
summary_form_rework - the Chief Reviewer gives the completed Review Summary Form to the Release Team
questions - Reviewers raise and discuss questions with the Author during the review meeting
defects - Reviewers raise and discuss defects with the Author during the review meeting
discussion - Reviewers and the Author discuss various issues related to the line item during the review meeting
re-review_decision - all Reviewers decide whether or not a re-review is required
rework - the Author, Line Item Owner, and Chief Reviewer review the rework

Not all of the interactions listed above occurred during all reviews. Those that did occur, however, are represented just once for that review. For example, there is only one **questions** interaction for each review meeting. Although a number of questions may have been raised and discussed, they all involve the same set of people, use the same communication media, and refer to the same document (the design or code being reviewed). Thus all of the independent variables have the same values for each question raised. Consequently, for notational and computational convenience, we have modeled the **questions** interaction as a single interaction per review. The same is true for the **defects** and **discussion** interactions.

These identified interactions constitute the unit of analysis for this study, as explained in section 3.2. That is, each data point corresponds to an interaction of one of the types listed above. In addition, the *type* of an interaction (e.g. **defects**, **rework**, etc.) is one of the variables we shall use in the analysis of the data.

3.4. Data Collection

The data for this study were collected during an initial visit to IBM in Toronto in June 1994, two follow-up visits in November 1994 and April 1995, and several email communications in between the visits. The data collection procedures included gathering of official documents, participant observation [Taylor84], and structured interviews [Lincoln85]. The observations and interviews were guided by a number of forms, or instruments. All of these procedures and instruments are discussed in the following sections.

Checklist for observing reviews:

Release:
Line Item:
Review type:
Date:
Author(s):

Moderator:
Chief Reviewer:
Reviewers:

Total prep:
Meeting length:
 # participants:
Amount reviewed:
Is this a re-review?
Is a re-review planned?

Defects:	**FPFS**	**HLD**	**LLD**	**CODE**	**TPLAN**	**TCASES**
Major						
Minor						

Time to:
read (1 person on average):
fill out summary form:
questions:
discuss errors:
other discussion:

Log (on other side)
Categories of time:
Questions (Q)
Error discussion (E)
Other discussion (D)
Summary form (SUM)
Administration (A)

Notes:

Figure 4. The observation checklist

3.4.1. Documents

The official documents of an organization are valuable sources of information because they are relatively available, stable, rich, and non-reactive, at least in comparison to human data sources [Lincoln85]. The model of the DB2 environment (described in section 3.3) relied initially on two confidential IBM documents, a review process document and the official organization chart.
Other documents which provided data later were copies of the *Review Summary Forms* for each review that was observed. Most of the information on these forms had already been collected during the observations of the reviews, so the forms served as a validation (triangulation) instrument.

3.4.2. Observations

Participant observation, as defined in [Taylor84], refers to "research that involves social interaction between the researcher and informants in the milieu of the latter, during which data are systematically and unobtrusively collected." Examples of studies based on participant observation are found in [Barley90, Perry94, Sandelowski92, Sullivan85]. In these examples, observations were conducted in the subjects' workplaces, homes, and therapists' offices. The idea, in all these cases, was to capture firsthand behaviors and interactions that might not be noticed otherwise.

Much of the data for this study was collected during direct observation of 10 reviews of DB2 line items in June 1994. Figure 4 shows a copy of the form, called the observation checklist, that was filled out by the observer for each review. Most of the administrative information on the form was provided in the announcement of the review, the Review Summary Form, or by the participants during or after the review meeting.

During the course of the review, each separate discussion was timed. The beginning and ending times, the participants, and the type of each discussion were listed on the back of the observation checklist for the review. A discussion constituted the raising of a defect (E) if it ended in the Moderator making an entry on a Major Defect Form or the list of defects he or she was keeping. A question (Q) was a discussion that did not end in an entry on the defect list, and additionally had begun with a question from one of the Reviewers. Other discussions (D) were those that neither began with a question nor ended with the recording of a defect. Some time was spent filling out the summary form (SUM), for example the reporting of preparation time for each Reviewer. Finally, a small amount of time in each review was spent in administrative tasks (A), for example connecting with remote Reviewers via phone. Totals for these various categories of time were recorded on the observation checklist.

3.4.3. Interviews

In [Lincoln85], a structured, or "focused", interview is described as one in which "the questions are in the hands of the interviewer and the response rests with the interviewee", as opposed to an unstructured interview in which the interviewee is the source of both questions and answers. The interviews conducted for the pilot study were structured interviews because each interview started with a specific set of questions, the answers to which were the objective of the interview. Examples of studies based on interviews can be found in [Rank92, Schilit82, Schneider85].

Interview Guide for John Doe, MM/DD/YY

1. How much of your prep time is spent filling out the defect form?
   Recording minor defects?

2. How much time is taken up by scheduling and distributing materials?

3. How long was the followup meeting?

4. Do you work much with the other participants, aside from reviews?

Figure 5. Example interview guide.

The initial interviews were conducted within a few days of each DB2 review. Other interviews took place in November 1994, and April 1995. One goal of each interview was to elicit information about interactions that were part of the review process but that took place outside the review meeting (and thus were not observed). The interviews also served to clarify some interactions that went on during the meeting, and to triangulate data that had been collected during the meeting. Before each interview, the interviewer constructed an interview form, or guide [Taylor84], which included questions meant to elicit the information sought in that interview. These forms were not shown to the interviewee, but were used as a guide and for recording answers and comments. Figure 5 shows an example of such a guide. For each review, at least one Author and, with one exception, the Chief Reviewer was interviewed. As well, in most cases, several other Reviewers were interviewed.

3.5. Measures

This section describes the procedures used to transform the data collected, as described in the last section, into quantitative variables. First the list of variables is presented, then the details of how the information from documents, observations, and interviews is coded to evaluate these variables.

3.5.1. Variables

The variables chosen for analysis, listed in Table 1, fall into three categories. First is the dependent variable, *Communication Effort.* Secondly, there is a set of independent variables which represent the issues of interest for this study, i.e., organizational structure. Several different measures have been chosen which capture different relevant properties of organizational structure. Finally, there is a large set of variables which are believed to have an effect on communication effort, but which are not the primary concern of this study. If these variables are not taken into account, they threaten to confound the results by hiding the effects of the organizational structure variables.

The dependent variable, labelled **CE** (for Communication Effort), is the amount of effort, in person-minutes, expended to complete an interaction. This is a positive, continuous, ratio-scaled variable and is calculated for each interaction.

There are four organizational structure variables. They are measured in terms of the set of participants in an interaction. The first two, **XOD** and **MOD,** are closely related. They are both based on *Organizational Distance*, which quantifies the degree of management structure between two members of the organization. Using a graph as shown in Figure 1, the shortest path between each pair of interaction participants is calculated. If a shortest path value is 4 or less, then this value is the Organizational Distance between that pair of participants. If it is more than 4, then the Organizational Distance for the pair is 5.

Note that this definition of Organizational Distance does not assume that the management structure is strictly hierarchical. Choosing the shortest path between each pair of participants allows for the possibility that more than one path exists, as might be the case in a matrix organization. It should also be noted that the links representing management relationships are not weighted in any way. One enhancement of this measure would be the addition of weights to differentiate different types of reporting relationships.

The higher values of Organizational Distance have been combined into one category for two reasons. First of all, the data showed that most pairs of interaction participants had an Organizational Distance of 4 or less. Also, it was impossible to calculate Organizational Distance accurately between some very distant pairs of participants. For example, reviews sometimes included a Reviewer from another IBM company. In these cases, the management links to the outside Reviewer were not well defined. Any pair that included that participant would then have an Organizational Distance of 5.

XOD and **MOD** are both aggregate measures of Organizational Distance. They differ in the way that Organizational Distance values for individual pairs of participants are aggregated into a value for an entire set of participants. **XOD** is defined as the maximum Organizational Distance, and **MOD** is the median Organizational Distance, among all pairs of participants in an interaction. Therefore, **XOD** would be high for those interactions in which even just one participant is organizationally distant from the others. **MOD** would be high only for those interactions in which at many of the participants are organizationally distant. The median was chosen for **MOD** because the shortest path values for each pair of participants are ordinal, not interval, and so the mean is not appropriate.

The two other organizational variables are *Familiarity* (**Fam**) and *Physical Distance* (**Phys**). They are also based on pairs of interaction participants, and rely on graphs like those shown in Figures 2 and 3. They are both ordinal. Their levels are shown in Table 1. Familiarity reflects the degree to which the participants in an interaction work together or have worked together outside the review, and thus presumably share common internal representations of the work being done. The familiarity measure also attempts to capture the important informal networks. Physical Distance reflects the number of physical boundaries (walls, buildings, cities) between the interaction participants.

Variable Name	Label	Levels: Code	Levels: Meaning	Data Source(s)
Communication Effort (ratio)	CE	-	amount of effort, in person-minutes, expended to complete the interaction	observations; 6/94 and 5/95 interviews; Review Summary Forms
Organizational Distance (ordinal)	MOD	1-4 5	value of the median shortest path between participants in management structure if it is between 1 and 4 value of the median shortest path between participants in management structure if it is greater than 4	organization charts; observations; 6/94 and 5/95 interviews
	XOD	1-4 5	value of the maximum shortest path between participants in management structure if it is between 1 and 4 value of the maximum shortest path between participants in management structure if it is greater than 4	organization charts; observations; 6/94 and 5/95 interviews
Familiarity (ordinal)	Fam	1 2 3 4	pairs of participants who are familiar with each others' work <= 10% 10% < pairs of participants who are familiar with each others' work <= 20% 20% < pairs of participants who are familiar with each others' work <= 50% pairs of participants who are familiar with each others' work > 50%	6/94 and 5/95 interviews
Physical Distance (ordinal)	Phys	1 2 3 4	all participants in same office all participants on same corridor, but not all in same office all participants in Toronto but not all on same corridor at least one pair of participants at different sites	observations; online directory; 5/95 interviews
Number of Participants (absolute)	N	-	number of people participating in an interaction	observations; Review Summary Forms
Skill Level (ordinal)	K	1 2 3	low medium high	11/94 interviews
Request Medium (nominal)	Mr	0 1 2	no request made verbal request electronic request	observation; 5/95 interviews
Preparation Medium (nominal)	Mp	1 2 3 4 5	a written, paper form verbal, with no notes shared a brief written message a structured written document an unstructured written document	observation; 5/95 interviews
Transfer Medium (nominal)	Mt	1 2 3 4 5 6	face-to-face meeting conference call video conference electronic transfer paper 2-way phone call	observation; 5/95 interviews

Variable Name	Label	Levels		Data Source(s)
		Code	Meaning	
Information size (ordinal)	Size	1 2 3 4 5	very small <= 3 pages < 1 KLOC or 4-9 pages 1-2 KLOC or 10-25 pages > 2 KLOC or > 25 pages	observation; 6/94 interviews; Review Summary Forms
Technicality (ordinal)	Tech	1 2 3	non-technical mixed technical	dictated by type of interaction
Complexity (ordinal)	Comp	1 2 3 4 5	very easy easier than average average more difficult than average very difficult	6/94 interviews
Structure (ordinal)	Struct	1 2 3	highly structured mixed unstructured	observation
Information Use (nominal)	Use	1 2 3 4	informational decisional directional functional	dictated by type of interaction
Interaction Type	Type		see Section 3.3.3	

Table 1. Variables used in this study

The set of blocking variables is large. The first is the size of the set of interaction participants. This variable, labelled **N**, is simply the number of people who expend effort in an interaction.

Another blocking variable is skill level, **K**. This variable reflects the level of skill possessed by the person most responsible for an activity, relative to the skills required to complete the activity. The assumption is that the more skilled a person is, the less he or she will depend on other people, and consequently the less time he or she will spend in communication. Many very simple interactions have a high (3) value for **K**, because such interactions do not require much skill. For other interactions, which are more technical in nature, **K** is set equal to the skill level of the Chief Reviewer.

We also wish to block according to the type of communication media used in an interaction. Three different parts of an interaction have been identified that require (potentially different) communication media. The first is the medium used to request information. Since many interactions involve unsolicited information, there is no request and thus no medium for this purpose. In the interactions we studied, when such a request was made, it was made either verbally or via email. Thus this variable, labelled **Mr**, is coded as either a 0 (n/a), 1 (verbal), or 2 (email).

The second part of an interaction which is affected by the choice of communication medium is the preparation of the information to be shared. In the interactions studied, information was prepared in one of five different ways. Some information required simply the completion of a paper form. Other information was to be shared verbally, and required only that the sender prepare his or her thoughts. In this case, written notes might be prepared, but were not shared with other participants in the interaction. The third preparation medium is the writing of a brief, informal message. Fourth is the preparation of a formal document which follows a defined format and structure. Finally, some interactions required the information in the form of an unstructured document. The blocking variable **Mp**, the medium used to prepare the information to be shared, is coded as a 1 (form), 2 (verbal), 3 (message), 4 (structured document), or 5 (unstructured document).

Finally, an interaction requires a communication medium to transfer the information between participants. The "transfer" media that were used in the interactions studied were face-to-face meetings, conference calls, video conferences, electronic transfer (email, ftp, etc.), paper, and normal phone calls. These values of the variable, labelled **Mt** are coded with numbers 1 through 6, respectively.

The last few blocking variables concern properties of the information that is to be shared in an interaction. The first of these variables is the amount of information. The information in most interactions studied was represented by the material that was being reviewed. In some cases, this material was code, measured in LOC, and in others it was a design document, measured in pages. In order to collapse these two "size" measures, we have created an ordinal scale, shown in Table 1, under "Information Size". The first level refers to interactions where the information shared is very simple, e.g. the answer to a "yes or no" question. These interactions are normally part of the managerial tasks surrounding a review. The second level is also used in some managerial interactions that involve a bit more information, e.g. a review summary form, and also reviews of small design documents. The last three levels each correspond to both a number of lines of code and a number of pages. The two definitions of each level were considered roughly equivalent by a number of experienced reviewers. The boundaries between these levels were chosen based on the data, which naturally fell into these groups. The amount of information in an interaction is labelled **Size**.

The second information characteristic is the degree of technicality of the information. This variable, labelled **Tech**, is coded as 1 for non-technical (managerial or logistical) information, 2 for information which is mixed, and 3 for purely technical information.

Information complexity, **Comp**, is coded on a five-point subjective scale, based on the answers to questions put to developers about the comparative complexity of the material being reviewed. **Comp** ranges from very easy (1) to very hard (5). This variable is meant to capture how much difficulty interaction participants would have in understanding the information. Information in managerial or logistical interactions is generally not very complex (usually 1). Review materials, however, vary over the entire range.

The degree of structure that the information exhibits, labelled **Struct**, reflects whether or not the information follows some predefined format or is in some language more precise than English. Source code, for example, is highly structured (1), as is information on a form. Questions and discussion are unstructured (3). Design documents, which are written with a predefined template, are in between (2).

The use to which it is to be put after the interaction is another characteristic of information. That is, we want to record the purpose of the interaction and the reason the information needs to be shared. This variable also gives some indication of the *importance* of the information. This variable, labelled **Use**, has the value 1 if the purpose of the interaction is general information, and it is not clear what specific activities or decisions will be affected by this information in the future. A value of 2 indicates that the information will be used to help make a decision. Some information is used to influence how and which activities are performed (3). This is often logistical information, for example a deadline. Finally, information can be used as input to an activity (4), for example a design document as input to the implementation activity.

Finally, we will use the type of interaction, **Type**, as a variable later in our analysis. The type of an interaction is related to the step of the process it is part of and the information involved. The types of interactions for this study were identified during construction of the three-part model presented in section 3.3. The list of interaction types is presented in section 3.3.3.

3.5.2. Coding

Recall that each review has associated with it a set of interactions, each of which has a value for every variable. Thus, there is an instance of each variable for each interaction in each review. The values of some independent variables are completely determined by the type of interaction. That is, some variables have the same value for every interaction of a certain type, regardless of which review it is part of. The values of these variables are dictated by the way that interactions have been modeled. For other variables and other interactions, the values vary over different reviews. This information is summarized in Table 1.

The dependent variable, **CE**, was coded by combining several pieces of data, depending on the interaction. Recall that **CE** for an interaction is defined as the total amount of effort expended on that interaction, from the initial request for information through the reading (or otherwise digesting) of the information.

The effort for many interaction types (e.g. **schedule_meeting**, **choose_participants** and **preparation_done**) is straightforwardly gathered from interviews. In many cases, the effort information gathered in interviews reflects the amount of time just one of the participants spent in the interaction, so the value must be multiplied by the number of participants.

The effort for the interactions which take place during the review meeting is actually observed and timed. These values must also be multiplied by the number of people present at the meeting. This is not always straightforward, as it was common for reviewers to come and go during the course of the review meeting. Some of these interactions also include some of the preparation time (e.g. reviewers prepare the defects to be raised at the meeting ahead of time), so that is included in the calculation of **CE**.

Values of the organizational variables, Organizational Distance (**XOD** and **MOD**), Familiarity, and Physical Distance, are calculated directly from the graphs that make up the organizational part of the model described in section 3.3.2. The scales used for these variables are shown in Table 1. These scales were derived from the data itself.

Most of the information used to evaluate the blocking variables was collected during interviews. Some blocking variables, however, were evaluated a priori according to the type of interaction. For example, some interaction types always involve technical information while others are concerned with purely managerial or logistical information. So the value of the variable **T** is constant for each type of interaction, regardless of any characteristic of individual reviews.

3.6. Data Analysis

The data to be analyzed consisted of 100 data points, each corresponding to a single interaction. Associated with each data point were values for each of the independent,

dependent, and blocking variables. In addition, we recorded for each data point what type of interaction it corresponded to, which review that interaction was a part of, and the nature of the material being reviewed (code, design, etc.).

The data analysis involved the construction of histograms to display the distributions of individual variables, and scatterplots to illustrate the relationships between pairs of variables. Blocking variables were used in a limited way to explore whether or not relationships between variables held under different conditions. Part of the analysis also involved blocking the data by interaction type.

Our analysis method basically consisted of creating subsets of data based on the values of one or more variables, then creating histograms and scatterplots based on those subsets. The subsets of interactions that we analyzed are:

- the entire set of interactions
- high effort interactions (**CE**>250 and **CE**>500)
- technical interactions which take place during the review meeting (**questions**, **defects**, and **discussion**).
- by technicality
- by complexity
- by degree of structure
- by size of information
- by skill level
- by number of participants
- by interaction type
- by line item
- by combinations of the organizational variables (e.g. low Physical Distance and high **MOD**)

For each of these subsets, a histogram showing the distribution of **CE** was generated, as well as scatterplots showing the relationships between **CE** and each organizational variable (**MOD**, **XOD**, Physical Distance, and Familiarity). To test these relationships, Spearman correlation coefficients were calculated. In addition, the distributions of other variables were analyzed for some of the subsets. For example, we looked at the distribution of interaction types among the high-effort interactions. Also, for both the data set as a whole and for the high-effort interactions, we studied the distributions of all variables. Another two-variable relationships that we explored with scatterplots is the relationship between **CE** and the number of participants (**N**). For this relationship, we grouped the data by line item to see which line items required more Communication Effort overall, and which required more effort per participant. We also ran ANOVAs (Analysis of Variance tests) on some combinations of variables for some subsets, but there was not enough data to yield meaningful results. Mann-Whitney tests were also used to test some special hypotheses about combined effects of Organizational Distance. The strongest and most interesting of our findings are presented in the next section.

4. Results

The results of our study are presented in four subsections. First, we give an overall characterization of the data collected by looking at each variable in isolation. Then we present some findings concerning the relationships between the dependent variable (Communication Effort) and the various organizational independent variables. We call these relationships "global" because they hold in the data set as a whole. In section 4.3, we examine those interactions that required the most Communication Effort more closely. Finally, in section 4.4, we divide the data by type of interaction to see what patterns emerge for different types.

4.1 Data Characterization

We begin by characterizing the data collected. In particular, we will examine the distributions of values of the variables studied. First, as can be seen in Figure 6, the distribution of the dependent variable, Communication Effort, is highly skewed towards the low end. The box plot at the top shows another view of this distribution. The box itself is bounded by the 25th and 75th quantiles. The diamond indicates the mean of the data. 90% of the interactions had a **CE** of less than 600 person-minutes.The maximum amount of effort any interaction required was 1919 person-minutes, and the minimum was 3 person-minutes. The median was 38 and the mean was about 190.

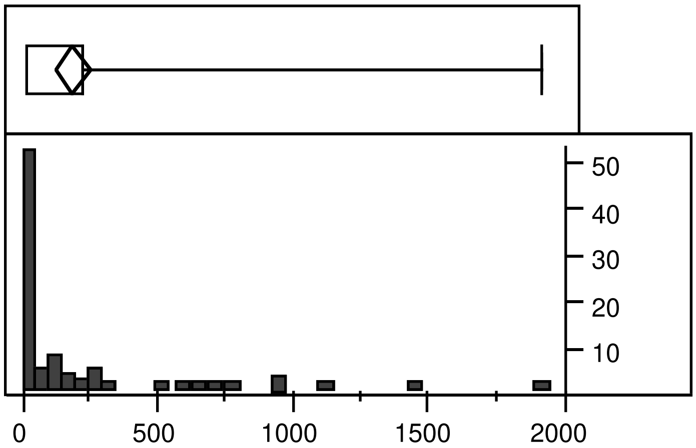

Figure 6. Distribution of Communication Effort (in person-minutes) over all 100 interactions

	MOD		XOD	
Level	Count	Cum %	Count	Cum %
1	32	32	23	23
2	29	61	1	24
3	6	67	0	24
4	27	94	33	57
5	6	100	43	100

Table 2. Frequency table for median (MOD) and maximum (XOD) Organizational Distance

It is also useful to look at the distribution of the independent variables. Table 2 shows the numbers and cumulative percentages of data points at each level of **MOD** and **XOD** (recall that there are exactly 100 data points, so simple percentages are not shown). About 60% of the interactions had a median Organizational Distance (**MOD**) of 2 or less, while more than three quarters had a maximum Organizational Distance (**XOD**) of 4 or higher. If we look at **MOD** and **XOD** together, as in Table 3, we see that most of the data falls into three categories:

- 24% of the interactions have all participants organizationally close (low **MOD**, low **XOD**)
- 37% of the interactions have most of the participants organizationally close, but a few organizationally distant (low **MOD**, high **XOD**)
- 33% of the interactions have most of the participants organizationally distant (high **MOD**, high **XOD**)

We will be referring to these categories later as we examine the differences between them.

MOD \ XOD	1	2	4	5
1	23	0	8	1
2	0	1	16	12
3	0	0	0	6
4	0	0	9	18
5	0	0	0	6

Table 3. Frequency of values for median and maximum Organizational Distance

Figures 7 and 8 show the distributions of Familiarity and Physical Distance. The interactions tend to have low Familiarity (75% with 2 or less) and high Physical Distance (83% with 3 or more).

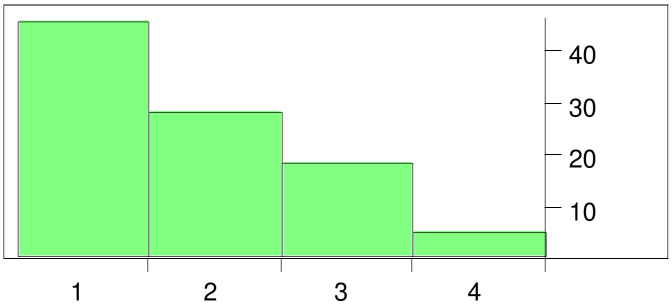

Figure 7. Distribution of Familiarity over all 100 interactions

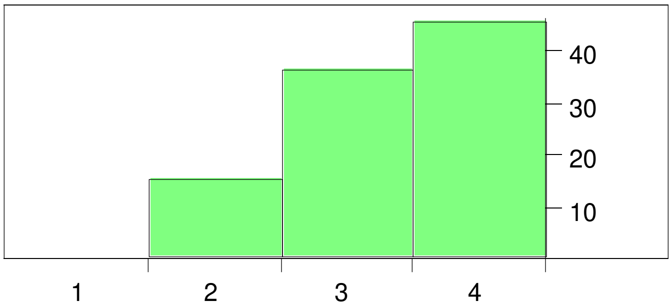

Figure 8. Distribution of Physical Distance over all 100 interactions

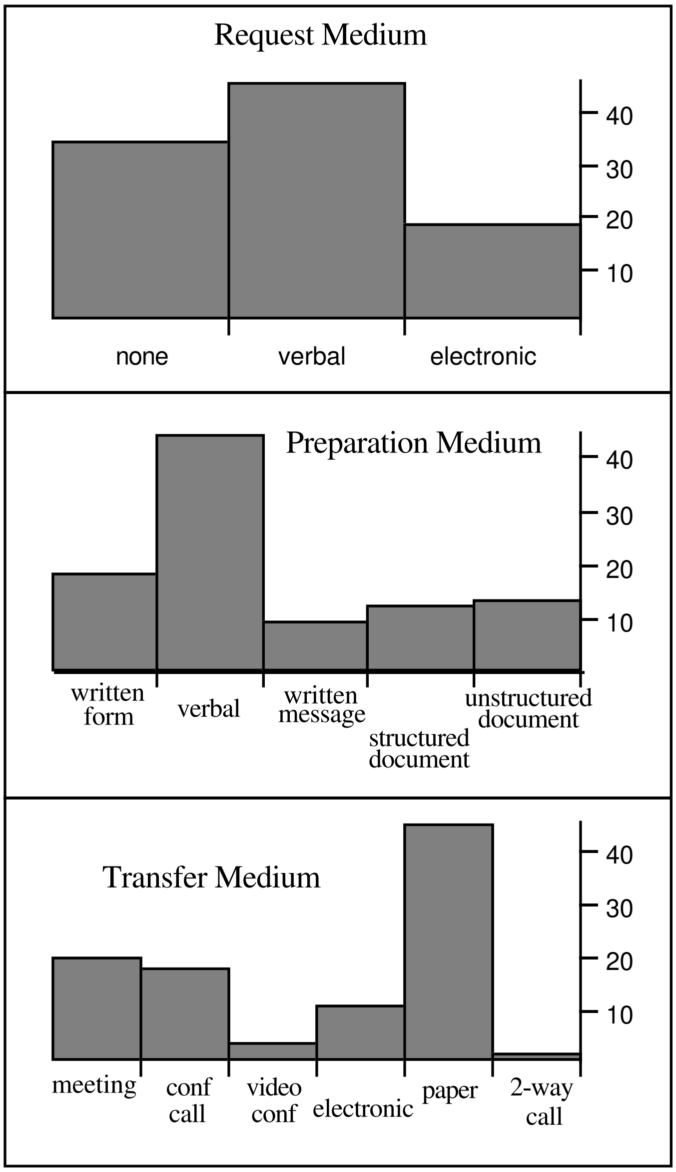

Figure 9. The different communication media used in all 100 interactions.

It is also useful to take a quick look at the characteristics of the data set with respect to the blocking variables. The distributions of the communication media variables are shown in Figure 9. Most interactions either began with a verbal request (46%), or no request at all (35%). The distribution of **Mp** shows that, in many interactions (44%), the information was prepared to be shared verbally. However, each of the other types of information preparation (written forms, messages, and documents) were used in 10-20% of the interactions. The medium most commonly used to actually transfer the information was paper (45%), although face-to-face meetings (20%) and conference calls (18%) were also well represented, along with email (11%).

All of the interactions involved either technical or non-technical information (no mixed), with about 60% of them technical. The information in most of the interactions (53%) was considered less complex than average, although a third were considered slightly more complex than average. The data was fairly evenly divided among interactions involving structured, unstructured, and mixed information. All different sizes of information were represented, although 50% of the interactions involved small amounts of information (3 pages or less). Almost half of the interactions involved information of a functional nature (**U**=4), 19% was directional, 5% was decisional, and 29% was informational. These distributions are shown in Figure 10.

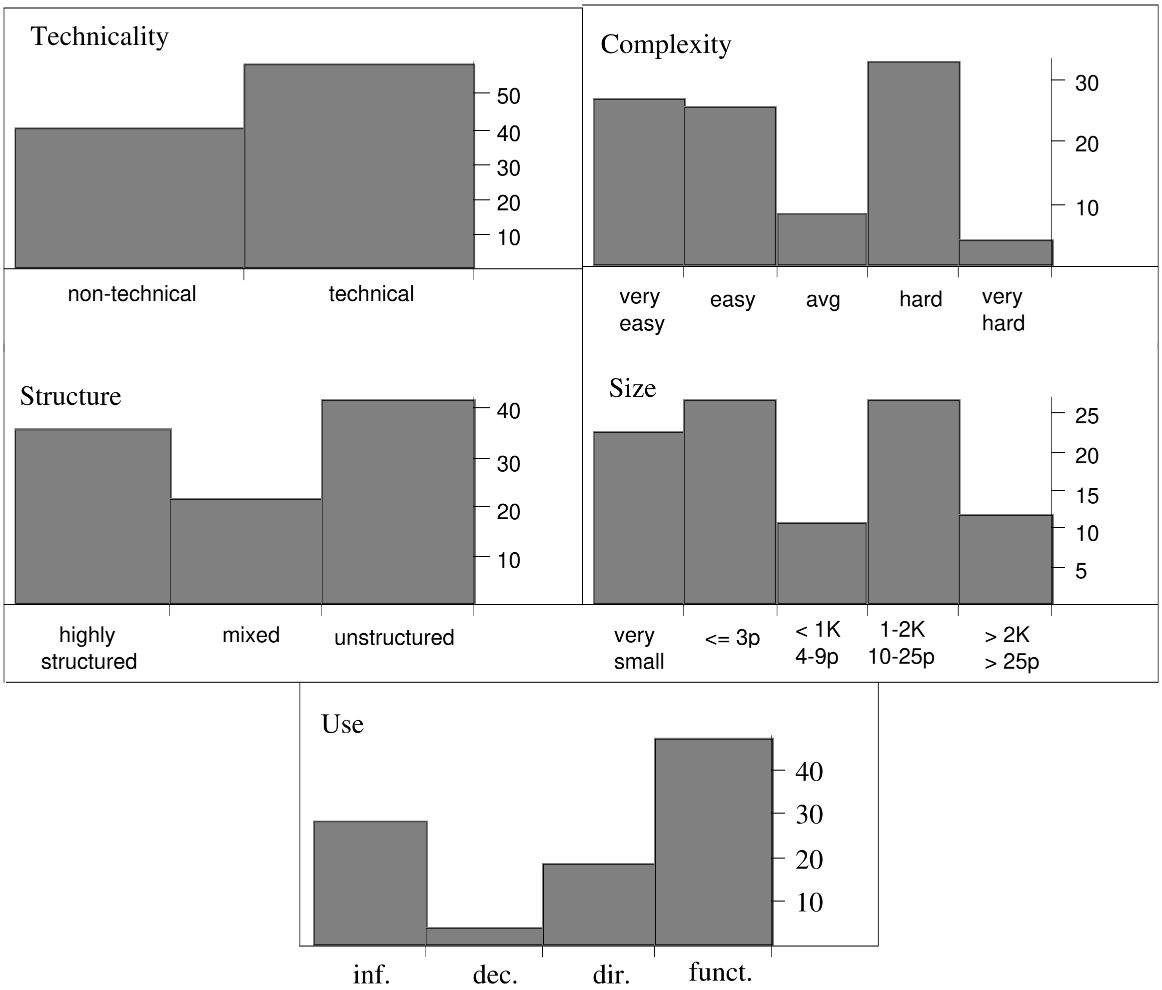

Figure 10. Distributions of blocking variables over all 100 interactions.

4.2 Global Relationships

Next, we want to look at the overall relationship between the dependent variable, Communication Effort, and each of the independent variables in turn. Figure 11 shows two scatterplots, each with Communication Effort on the vertical axis, and one of the two versions of the Organizational Distance variable on the horizontal axis. A boxplot is also shown for each level of each independent variable. The top and bottom boundaries of the boxes indicate the 75th and 25th quantiles. The median and the 90th and 10th quantiles are also shown as short horizontal lines (the median and 10th quantiles are not really visible on most boxes). The width of each box reflects the number of data points in that level.

From Figure 11, we can observe that the highest-effort interactions are those with a relatively low median Organizational Distance (**MOD**) and relatively high maximum Organizational Distance (**XOD**). This is the second category described above, in the discussion of the distributions of **MOD** and **XOD**. This observation implies that groups require more effort to communicate when they include a few (but not too many) members who are organizationally distant from the others. Less effort is required when the group is composed of all organizationally close members (low **MOD** and low **XOD**), or all or nearly all organizationally distant members (high **MOD** and high **XOD**). We tested the statistical strength of this result by calculating the Mann-Whitney U statistic. This is a non-parametric test meant to indicate whether or not two independent samples exhibit the same distribution with respect to the dependent variable (**CE**). In this case, the two groups were those interactions falling into the high **XOD**/low **MOD** category, and those which did not. The test yielded a significant value, even at the p<.01 significance level.

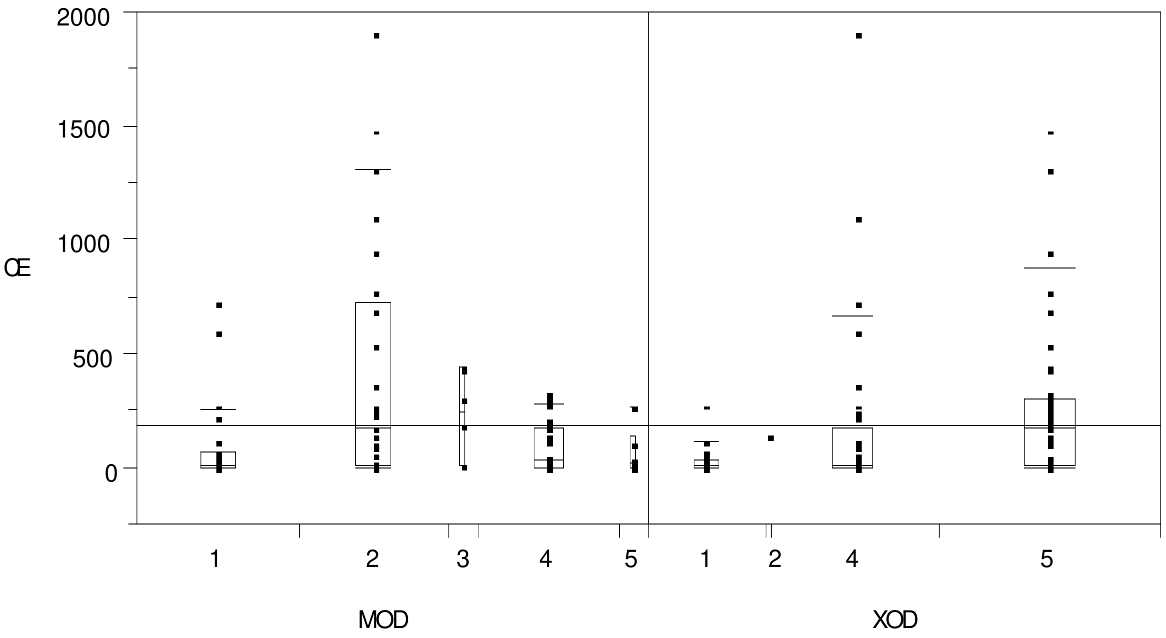

Figure 11. Communication Effort plotted against median Organizational Distance (MOD) and maximum Organizational Distance (XOD)

We investigated this interesting result about Organizational Distance in more detail by partitioning the data by values of the different blocking variables, and then performing the same Mann-Whitney test on each partiition. Again, this test was performed to determine if interactions in the high **XOD**/low **MOD** category exhibited significantly higher levels of **CE** than other interactions. The test was run using both normalized and unnormalized **CE** values for the dependent variable. The results are summarized in Table 4. There are some values of some independent variables that are not used to restrict the data set because the resulting subsets were too small or too homogeneous to yield meaningful results. The values in the table are the "p" values, which indicate, in each case, the probability that the difference in **CE** between interactions with high **XOD**/low **MOD** and other interactions is due to chance. In other words, a low value for p indicates a significant difference in **CE**. Generally, a value of .05 or less is considered significant.

Data set restricted to those interactions with:	Probability that the difference in **unnormalized CE** is due to chance	Probability that the difference in **normalized CE** is due to chance
Skill Level = medium	**0.0018**	**0.03**
Skill Level = high	**0.02**	0.83
Use = informational	**0.02**	**0.07**
Use = directional	0.13	0.74
Use = functional	0.33	0.96
Size = very small	0.59	0.4
Size <= 3 pages	0.79	0.79
Size = <1KLOC or 4-9p	0.24	0.64
Size = 1-2KLOC or 10-25p	**0.04**	0.06
Size > 2KLOC or 25pages	**0.01**	**0.03**
Structure = highly	0.39	0.33
Structure = mixed	0.84	0.25
Structure = unstructured	**0.004**	**0.03**
Complexity = very easy	**0.007**	0.29
Complexity = easy	0.87	0.87
Complexity = difficult	**0.001**	0.07
Technicality = non-technical	0.87	0.004
Technicality = technical	**0.002**	**0.04**
Mt = face to face	0.29	0.64
Mt = conference call	0.19	0.26
Mt = electronic	1.0	0.93
Mt = paper	0.1	0.1
Mp = paper form	1.0	1.0
Mp = verbal	0.22	0.28
Mp = written message	0.5	0.75
Mp = structured document	0.67	0.67
Mp = unstructured doc.	0.1	0.1
Mr = no request	**0.04**	0.32
Mr = verbal request	0.19	0.24
Mr = electronic request	0.11	0.17
Phys = 3	0.42	0.69
Phys = 4	0.09	0.58
Fam = 1	**0.0004**	0.45
Fam = 2	0.2	0.06
Fam = 3	**0.02***	0.06

Table 4. p values for the Mann-Whitney U test, comparing **CE** values of interactions with high **XOD**/low **MOD** with other interactions, with the data restricted by the values of other independent variables. Significant values are highlighted.
(*) When the data is restricted by **Fam** = 3, the difference in **CE** is significant but in the opposite direction than expected.

Surprisingly, the results of the test were not significant for many of the data subsets. It was significant for medium skill level, and for unnormalized **CE** at the high skill level. It was also significant, at least for unnormalized **CE,** for interactions involving large amounts of information (the two highest levels of **Size**) but not smaller amounts. The difference was significant for informational interactions, but not for directional or functional (levels of **Use**). Significance was found for interactions involving unstructured information, but not mixed or highly structured (**Struct**). Significance (in unnormalized **CE**) was also found for two different levels of complexity, "very easy" and "more difficult than average", but not the others. Significance also held for technical interactions, but not administrative ones. Significance was not found for any individual levels of any of the communication media variables, with one exception, nor for any levels of Physical Distance. Partitioning the data by levels of the Familiarity variable produced some interesting results. The Mann Whitney test found a significant difference in unnormalized **CE** for interactions with **Fam** equal to 1, but not 2. For interactions with **Fam** equal to 3, the test was significant, but in the opposite direction. That is, in this subset, interactions in the high **XOD**/low **MOD** category exhibited significantly *lower* levels of unnormalized **CE** than other interactions.

This set of results is difficult to interpret. In general, interactions which have high **XOD** and low **MOD** will require more communication effort. However, the effect of Organizational Distance may be overshadowed by the effect of size, use, degree of structure, complexity, or technicality.

In Figure 12, Communication Effort is plotted against the two remaining independent variables, Familiarity (**Fam**) and Physical Distance (**Phys**). It appears that high effort is associated with low Familiarity and with high Physical Distance (the latter observation being the strongest). However, it must be noted that most interactions have low Familiarity and high Physical Distance.

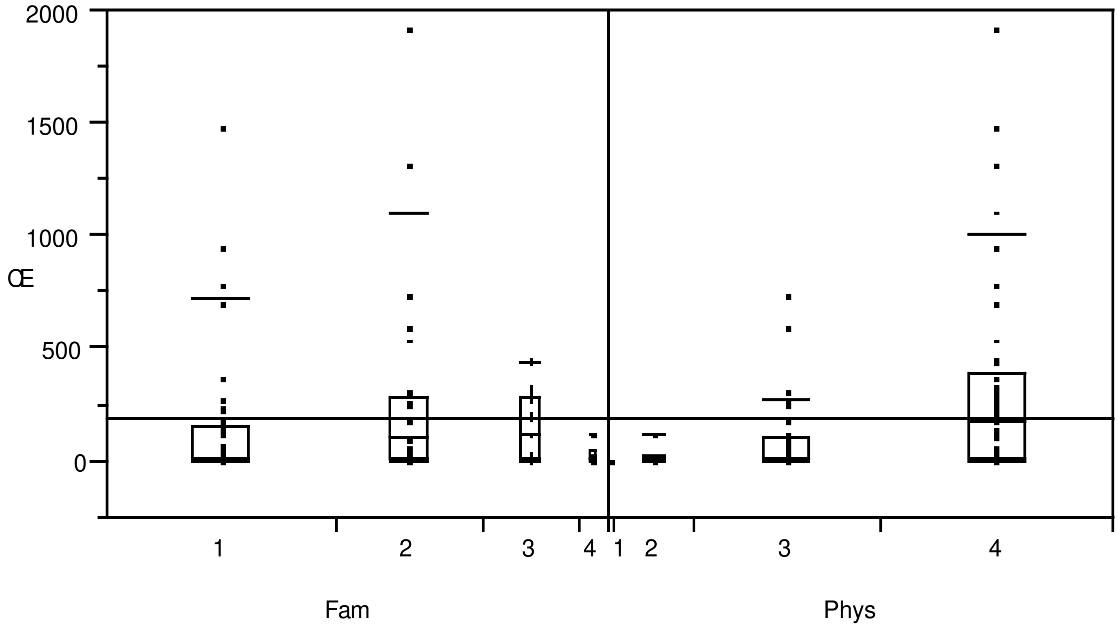

Figure 12. Communication Effort plotted against Familiarity (Fam) and Physical Distance (Phys)

The Spearman correlation coefficients, which reflect the strength of the relationships between each independent variable and the dependent variable, are shown in Table 5.

MOD	XOD	Fam	Phys
0.09	0.4	0.14	0.5

Table 5. Spearman rho (ρ) coefficients comparing each independent variable to the dependent variable, **CE** .

4.3 High Effort Interactions

In order to investigate all of these possible relationships, we have examined in more detail the subset of interactions which were effort-intensive. In particular, we have chosen the 11 highest-effort interactions, all of which required a Communication Effort greater than 500 person-minutes, and compared the characteristics of this subset to the distributions of the entire subset, described above. The first observation is that **CE** is more evenly distributed in this subset, as can be seen in Figure 13.

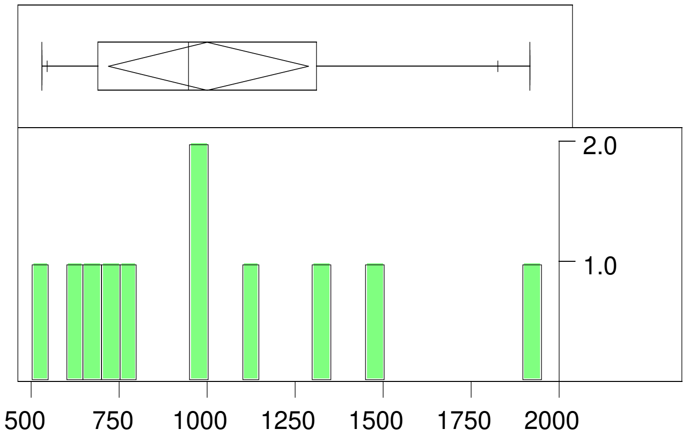

Figure 13. Distribution of Communication Effort over the 11 highest-effort interactions. The y axis is the number of data points.

It should also be noted that the product development team that was studied was divided into two subteams. Each subteam was developing a version of DB2 for a different hardware platform. All of the high-effort interactions took place during reviews conducted by just one of the teams. Most of the high-effort interactions were also either of type **defects** or **questions**.

MOD \ XOD	4	5
1	2	0
2	2	7

Table 6. Frequency of values for median and maximum Organizational Distance for the 11 highest-effort interactions

In looking at the distributions of Organizational Distance in this subset, we noticed that none of the high-effort interactions had a **MOD** more than 2, and none had a **XOD** less than 4. In fact, all of the interactions in this high-effort subset belong to the second category (low **MOD**/high **XOD**) described above, as shown in Table 6.

Also in this subset, we see the same pattern in the Familiarity and Physical Distance variables (Figure 14). That is, interactions tend to have low Familiarity and high Physical Distance both in the data set in general and in the high-effort subset. However, this trend is accentuated in the subset, where none of the interactions have Familiarity more than 2 (as compared to 25% in the whole data set). Similarly, 80% of the high-effort interactions have a Physical Distance of 4, the highest level of this variable.

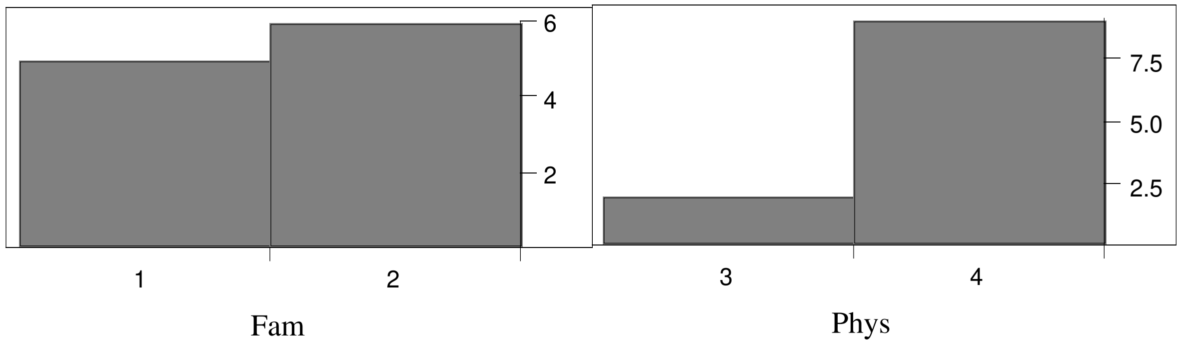

Figure 14. Distributions of Familiarity and Physical Distance among 11 highest effort interactions.

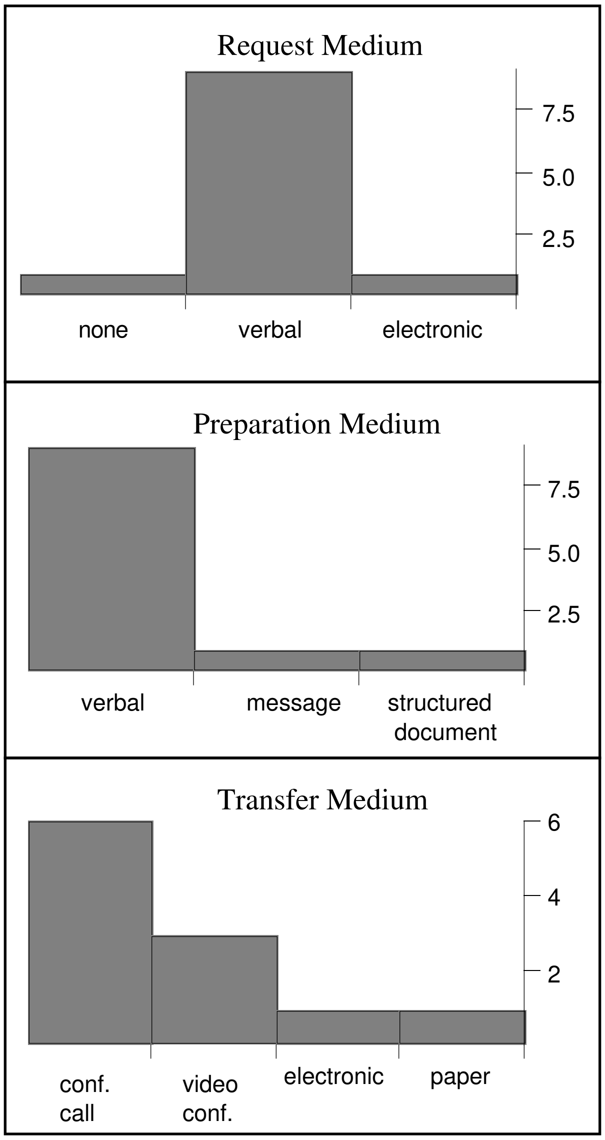

Figure 15. The different communication media used in the 11 highest-effort interactions.

Nearly all of the high-effort interactions involved a verbal request for information (**Mr**=1), no written preparation of the information (**Mp**=2), and were executed using a conference call or video conference (**Mt**=2 or 3). These patterns in the use of communication media, shown in Figure 15, differ dramatically from the patterns seen in the data as a whole. Interactions which involved a verbal request and no preparation usually took place during a face-to-face meeting in which many people were present, which implicitly increases the communication effort. In those meetings in which conference calling or videoconferencing was used, the technology actually slowed down the process. Significant amounts of time were spent waiting for remote participants to find the right page, to clarify issues for remote participants, etc. Also, the communication technology was unfamiliar to some participants.

All of the high-effort interactions involved technical information. This would imply that developers do not spend a large amount of time in administrative (non-technical) communication. In this study, 40% of all interactions were administrative in nature and none of them were highly effort-intensive. In fact, over all 10 reviews studied, 96% of the effort spent in communication involved technical information.

Comparisons between high-effort interactions and the whole set of interactions in terms of the other blocking variables yield few surprising results. The information in most of the high-effort interactions was unstructured, medium to large in size, and of average or higher complexity. The different uses of information in the high-effort interactions were not very different from that in the entire set of interactions, nor were the skill levels of the participants.

One other variable deserves a little more attention. The median number of participants in high-effort interactions is 10, but the median in the larger set of interactions is about half that (5.5). This result is not so straightforward as it might seem, however, because the variable **N** (number of participants) is not completely independent from Communication Effort. For some interactions, in fact, **N** is used in the calculation of **CE**. For example, **CE** for the interaction of type **discussion** is calculated by multiplying the amount of time spent in general discussion during the review meeting by **N**. To investigate whether or not the number of participants has an independent effect on effort, we normalized Communication Effort by dividing it by **N**. Then we picked the 15 interactions with the highest normalized **CE**(15 was the smallest number which included the 11 interactions we analyzed before as the highest-effort). The median number of participants in this subset is 8, lower than than 10, but still considerably higher than the median of the data as a whole (5.5). So it appears that the highest-effort interactions involve more participants than interactions in general, regardless of which way effort is calculated.

In some of the discussion below, we refer both to "normalized" and "unnormalized" values of Communication Effort (**CE**). **CE** values are normalized simply by dividing them by **N**, as in the discussion above.

4.4 Interaction Types

Many of the types of interactions (**defects**, **review_material**, etc.) in this study differ in character from each other. Some of our most interesting results have come from studying each interaction type in isolation. Table 7 shows all the interaction types and relevant statistics, sorted by mean Communication Effort (unnormalized). Statistics based on normalized (by number of participants) **CE** are also shown for each interaction type.

Interaction Type	N	Unnormalized Mean	S.D.	Min	Max	Normalized Mean	S.D.	Min	Max
defects	11	642	658	92	1919	67	68	18	240
questions	11	409	373	68	1109	43	37	17	139
review_material	11	313	164	120	729	35	26	16	104
discussion	9	195	241	12	694	18	20	2	53
schedule_meeting	9	105	188	20	600	14	24	1	75
comments	7	97	92	9	273	41	46	3	137
choose_participants	2	65	78	10	120	33	39	5	60
rework	2	30	0	30	30	15	0	15	15
commented_material	7	22	22	4	60	9	8.8	1	23
preparation_done	3	20	0	20	20	4	1	2	5
summary_form	10	13	10	3	32	6	5	2	16
sum_form_rework	9	11	2	10	15	6	1	5	8
re-review_decision	9	8	4	4	13	1	0	0	1

Table 7. Mean Communication Effort by type of interaction, both unnormalized and normalized by the number of participants

The **defects** interaction is, on average, the most effort-intensive interaction type, whether or not Communication Effort is normalized by the number of participants. This makes intuitive sense, since this interaction embodies the entire purpose of the review. In our data set, most of the **defects** interactions involved a set of participants that fell into the second category (low **MOD**/high **XOD**), including all of those with **CE** above the mean. All of the **defects** interactions were verbal, and took place during a face-to-face meeting, conference call, or videoconference. The highest-effort **defects** interactions took place during conference calls. **Defects** interactions included anywhere from 4 to 15 participants, with a mean of about 9 participants. All of the **defects** interactions with (normalized or unnormalized) **CE** above the mean had 7 or more participants.

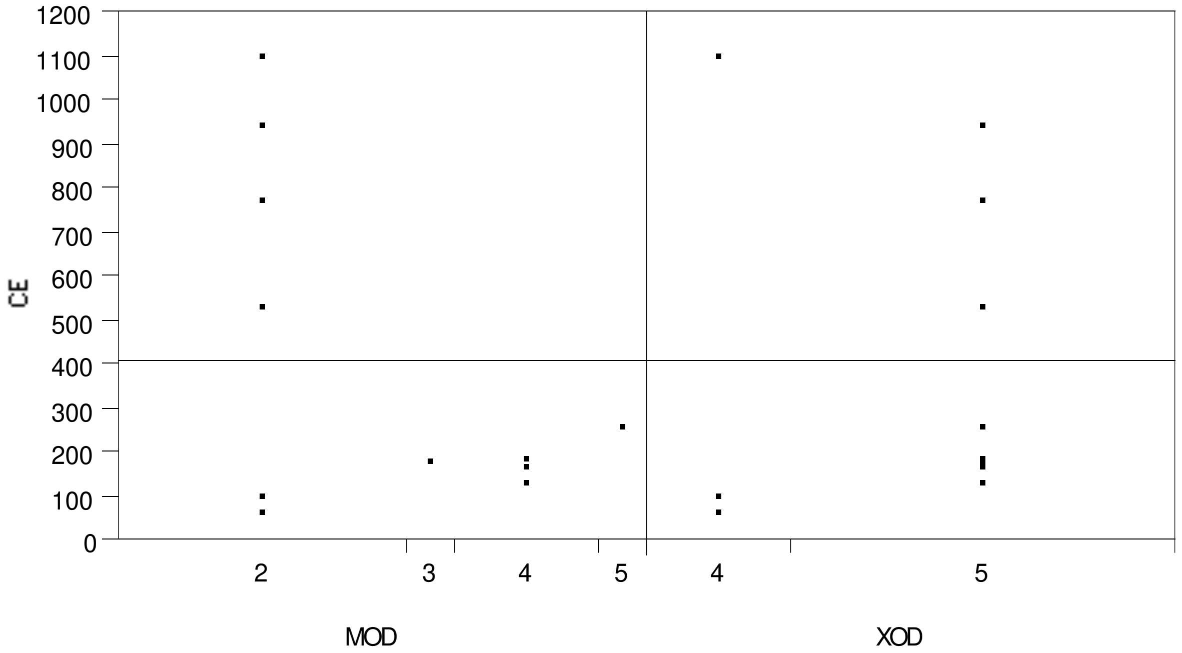

Figure 16. Communication Effort plotted against median and maximum Organizational Distance for **questions** interactions only

Another effort-intensive interaction type is the **questions** interaction. Again, the highest-effort interactions of this type fall into the second category of participant sets (low **MOD**/high **XOD**). This can be seen clearly in Figure 16.

Although all of the **questions** interactions have fairly high Physical Distance (3 or higher), the highest-effort ones have the highest Physical Distance, 4. Like the **defects** interactions, the above-average effort intensive **questions** interactions all had 7 or more participants.

The **discussion** interactions tend to be less effort-intensive than the **questions** or **defects** interactions, but still require more effort than most interaction types. **Discussion** interactions exhibit the same patterns in Organizational and Physical Distance as mentioned above for the **questions** and **defects** interactions. In addition, high-effort **discussion** interactions tend to involve information of relatively high complexity (**Comp**=4 or higher) and large size (**Size**>=4).

The **defects**, **questions**, and **discussion** interactions constitute all of the technical communication that takes place during a review meeting. The effort recorded for these interactions includes the effort required to prepare for, carry out, and digest this technical information. Since these interactions form the core of the work of a review, it is comforting to know that they are the ones which require the most effort. In fact, over all 10 reviews studied, 70% of the total Communication Effort expended was expended in interactions of these three types.

Two other relatively high-effort types of interactions, **review_material** and **comments**, exhibit slightly different behavior than described above. First of all, most of the high-effort **review_material** interactions have participants which fall into the third category described earlier (organizationally distant). The sets of participants for the high-effort **comments** interactions, on the other hand, fall into the first category (organizationally close). This contrasts with the observation that the participants in high-effort **defects**, **questions**, and **discussion** interactions are all in the second category. Another difference is that there is no apparent relationship between Communication Effort and Physical Distance for these types of interactions.

These results must be interpreted remembering that the set of participants in the **defects**, **questions**, and **discussion** interactions for each review is different than the set of participants for the **review_material** and **comments** interactions. The first three interactions take place during the review meeting, and the participants comprise all those present at the meeting. Furthermore, all the distance measures are calculated using every pair of participants. The distance measures for **review_material** interactions, on the other hand, reflect only the distances between the Author(s) and each Reviewer (i.e. not between Authors or between Reviewers). Thus, saying that the participants in a **review_material** interaction are organizationally distant means that the Authors are organizationally distant from the Reviewers. Similarly, the participants in a **comments** interaction are the Moderator and the Author(s). So saying that a **comments** interaction has a low Physical or Organizational Distance refers only to the distance between the Moderator and the Author(s).

The **review_material** and **comments** interactions also exhibited some relationships between effort and some of the blocking variables which did not seem relevant with other types of interactions. For instance, the high-effort **review_material** interactions all involved material that was highly structured and large, and took place during reviews in which the Chief Reviewer was highly skilled. High-effort **comments** interactions all involved information that was more complex than average.

5. Limitations of the study

The major limitation of this study is its size and scope. It examines only 10 reviews, during one month, in a single development project. The amount of data collected (100 data points), relative to the number of relevant variables, is too small to assess statistical significance of many of the findings or to generalize the results in any way.

The three-part model built at the beginning of the study (see section 3.3) was extremely useful throughout as a framework for organizing the data and for communicating with developers. However, it could have been more useful. In particular, handling the data associated with interactions was cumbersome and limited somewhat the analyses which could be done easily. Some automatic support for managing this part of the model (or even handling the data itself through the model), as well as a better notation, was needed.

Another lesson learned from this study was that the interactions, as defined, did not naturally fit the way the participants thought about the review process. This made collecting and validating the data very difficult. For example, the Reviewers' preparation time had to be divided over several different interactions in order to fit the model. Some of it was included in the Communication Effort for the **defects** interaction, some for the **questions** interaction, etc. During the interviews, we asked some Reviewers how they divided their preparation time. We used their responses as a guideline, but we cannot be sure that the percentages are accurate or consistent. Modeling more in accordance with the process as it is enacted, and at a slightly higher level of abstraction, would help eliminate doubts about the accuracy of the data.

The design of the research variables and their levels in the pilot study was based on expert opinion and the literature, but the process of designing these measures was not very formal or well-documented. A more rigorous qualitative analysis is needed to support the design choices. Such an "ahead-of-time" analysis is part of what is called *prior ethnography*, a technique from qualitative research methods.

During data collection, the follow-up interviews after the observed reviews were vitally important. However, they could have been combined into just one interview for each interviewee. Instead, the questions were spread over several interviews over a period of 10 months. This led to memory, personnel turnover, and discontinuity problems. A single interview, as shortly after the review as possible, is preferred.

The observations in this study were not as rigorous as they could have been. The single observer was not very familiar with the application domain, and this sometimes made it difficult to determine what type of discussions were taking place during observations. As well, no reliability techniques were employed, such as audio- or videotaping the reviews, or having a second observer. This would have ensured better accuracy of the data. Also related to data accuracy, there were some variables that had no triangulation [Lincoln85] source. That is, there was only one data source for these variables. It would be better, and should be possible, to have at least two sources for each piece of information collected.

During observations and interviews, some field notes were taken in addition to the information on the interview forms and observation checklists. However, this data was not extensive or reliable enough to be used as part of the data analysis. If more faithful notes had been kept, this qualitative data could have been used to help explain and interpret the quantitative results. The collection of useful anecdotes and quotes would also have been facilitated by making the interview questions more open-ended, that is, by relaxing the structuredness of the interviews a little.

One of the goals of this study was to serve as a pilot for a larger study begun recently. Although small, this pilot study was valuable in clarifying a number of issues related to how this subject is best studied. The limitations discussed above have been remedied in the design of the larger study, which is being conducted at NASA Goddard Space Flight Center. The main differences between this study and the pilot study described in this paper are its setting, size, and scope. Other differences arise as a result of remedying the limitations described above. The main goal of this larger study, as in the pilot study, is to learn how organizational structure characteristics affect the amount of effort expended on communication.

The setting for the larger study is the team developing AMPT, a mission planning tool. This project involves about 15-20 developers, and is the subject project for another experiment exploring a new software development process, Joint Application Development. As in the pilot study, we are studying the review process (called the inspection process at NASA) in particular. However, we expect to observe a much larger number of reviews (on the order of 30-50) over a longer period of time (3-6 months beginning December 1995).

6. Discussion and Summary

We have addressed the broad problem of organizational issues in software development by studying the amount of effort developers expend in certain types of communication. We have described an empirical study conducted to investigate the organizational factors that affect this effort. The research design combined quantitative and qualitative methods in an effort to be sensitive to uncertainty, but also to provide well-founded results. These methods include participant observation, interviewing, coding, graphical data displays, and simple statistical tests of significance.

Our findings are best summarized as a set of proposed hypotheses. The results of this study point to the validity of these hypotheses, but they are yet to be formally tested. Many of the methods and measures described in this paper may be used to do so. However, even as untested hypotheses, these findings provide important preliminary insight into issues of organizational structure and communication in software development:

H1 Interactions tend to require more effort when the participants are not previously familiar with each others' work. This is consistent with Krasner's [Krasner87] findings about "common internal representations".

H2 Interactions tend to require more effort when the participants work in physically distant locations. Curtis [Curtis88] and Allen [Allen85] have had similar findings.

H3 Interactions which take place during a meeting (a verbal request and an unprepared reply) tend to require more effort than other interactions.

H4 Interactions which involve some form of communication technology (conference calling and videoconferencing in this study) tend to require more effort.

H5 Non-technical (administrative) communication accounts for very little (less than 5%) of the overall communication effort.

H6 More participants tend to make interactions more effort-intensive, even when the effort is normalized by the number of participants.

H7 In interactions that take place in a meeting, more effort is required when the set of participants includes mostly organizationally close members, but with a few organizationally distant members. This contrasts with Curtis [Curtis88], who hypothesized that the relationship between organizational distance and communication ease is more straightforward.

H8 Preparing, distributing, and reading material tends to take more effort when all participants are organizationally distant, and when the material is highly structured and large.

H9 Writing and distributing comments to authors during the review meeting tends to take more effort when all of the participants are organizationally close, and when the material being reviewed is complex.

Recall that the problem to be addressed is the lack of knowledge about how to manage information flow in a software development organization and process. This study is not sufficient to solve this problem, but it is a first step. In section 1.1, several symptoms of this problem were described. The first is the difficulty of planning for communication costs. The findings of this study could be used to help in planning by pointing out characteristics which increase communication costs in reviews. For example, more than average time should be allowed for review meetings in which most of the participants are organizationally close, but a few are from distant parts of the organization. Alternatively, assignments could be made in such a way as to avoid such configurations of review participants.

The second symptom of the process information flow problem is that we do not know how to identify or solve communication problems as they arise. For example, if during the course of a project, developers are spending much more time preparing for reviews than planned, the findings above indicate that the problem may be that the participants are too organizationally distant, or that the material is too large. The problem might be solved by choosing reviewers who are closer, or by breaking the material to be reviewed into smaller pieces.

The third point raised in section 1.1 as a consequence of the research problem is that of learning from experience. This study represents a very small first step in building the experience necessary to effectively manage information flow in software development organizations. The next step for the authors is the larger empirical study described briefly in section 5. But there are several next logical steps in this line of research. No attempt has been made in this study to determine how communication effort affects software quality or development productivity. An understanding of this issue is necessary for effective management support. As well, this study does not address the issue of communication *quality*, only *quantity*. One cannot assume that the two are equivalent. Finally, there needs to be more work in the area of actually applying this new knowledge to the improvement of software development projects, and the mechanisms needed to achieve such improvement.

References

[Allen85] Thomas J. Allen. Managing the Flow of Technology. The MIT Press, 1985.

[Ballman94] Karla Ballman and Lawrence G. Votta. "Organizational Congestion in Large-Scale Software Development". In Proceedings of the 3rd International Conference on Software Process, pages 123-134, Reston, VA, October 1994.

[Barley90] Stephen R. Barley. "The Alignment of Technology and Structure through Roles and Networks". Administrative Science Quarterly, 35:61-103, 1990.

[Bradac94] Mark G. Bradac, Dewayne E. Perry, and Lawrence G. Votta. "Prototyping a Process Monitoring Experiment". IEEE Transactions on Software Engineering, 20(10):774-784, October 1994.

[Curtis88] Bill Curtis, Herb Krasner, and Neil Iscoe. "A Field Study of the Software Design Process for Large Systems". Communications of the ACM, 31(11), November 1988.

[Davenport90] Thomas H. Davenport and James E. Short. "The New Industrial Engineering: Information Technology and Business Process Redesign". Sloan Management Review, pages 11-27, Summer 1990.

[Ebadi84] Yar M. Ebadi and James M. Utterback. "The Effects of Communication on Technological Innovation". Management Science, 30(5):572-585, May 1984.

[Galbraith77] Jay R. Galbraith. Organization Design. Addison-Wesley, 1977.

[Gilgun92] Jane F. Gilgun. "Definitions, Methodologies, and Methods in Qualitative
Family Research". In Qualitative Methods in Family Research. Sage, 1992.

[Hammer90] Michael Hammer. "Reengineering Work: Don't Automate, Obliterate". Harvard Business Review, pages 104-112, July 1990.

[Krasner87] Herb Krasner, Bill Curtis, and Neil Iscoe. "Communication Breakdowns and Boundary Spanning Acitivities on Large Programming Projects". In Gary Olsen, Sylvia Sheppard, and Elliot Soloway, editors, Empirical Studies of Programmers, second workshop, chapter 4, pages 47-64. Ablex Publishing, New Jersey, 1987.

[Liker86] Jeffrey K. Liker and Walton M. Hancock. "Organizational Systems Barriers to Engineering Effectiveness". IEEE Transactions on Engineering Management, 33(2):82-91, May 1986.

[Lincoln85] Yvonna S. Lincoln and Egon G. Guba. Naturalistic Inquiry. Sage, 1985.

[March58] J.G. March and Herbert A. Simon. Organizations. John Wiley, New York, 1958.

[Mintzberg79] Henry Mintzberg. The Structuring of Organizations. Prentice-Hall, 1979.

[Perry94] Dewayne E. Perry, Nancy A. Staudenmayer, and Lawrence G. Votta. "People, Organizations, and Process Improvement". IEEE Software, July 1994.

[Rank92] Mark R. Rank. "The Blending of Qualitative and Quantitative Methods in Understanding Childbearing Among Welfare Recipients". In Qualitative Methods in Family Research, Sage, 1992.

[Sandelowski92] Margarete Sandelowski, Diane Holditch-Davis and Betty Glenn Harris. "Using Qualitative and Quantitative Methods: The Transition to Parenthood of Infertile Couples". In Qualitative Methods in Family Research, Sage, 1992.

[Schilit82] Warren Keith Schilit. "A Study of Upward Influence in Functional Strategic Decisions". PhD thesis, University of Maryland at College Park, 1982.

[Schneider85] Larissa A. Schneider. "Organizational Structure, Environmental Niches, and Public Relations: The Hage-Hull Typology of Organizations as a Predictor of Communication Behavior ". PhD thesis, University of Maryland at College Park, 1985.

[Stinchcombe90] Arthur L. Stinchcombe. Information and Organizations. University of California Press, 1990.

[Sullivan85] Matthew J. Sullivan. "Interaction in Multiple Family Group Therapy: A Process Study of Aftercare Treatment for Runaways". PhD thesis, University of Maryland at College Park, 1985.

[Taylor84] Steven J. Taylor and Robert Bogdan. Introduction to Qualitative Research Methods. John Wiley and Sons, 1984.