Communication and organization in software development: An empirical study

by C. B. Seaman
V. R. Basili

The empirical study described in this paper addresses the issue of communication among members of a software development organization. In particular, we have studied interactions between participants in a review process. The question of interest is whether or not organizational relationships among the review participants have an effect on the amount of communication effort expended. The study uses both quantitative and qualitative methods for data collection and analysis. These methods include participant observation, structured interviews, graphical data presentation, and nonparametric statistics. The results of this study indicate that several organizational factors do affect communication effort, but not always in a simple, straightforward way. Not surprisingly, people take less time to communicate when they are familiar with one another and when they work in close physical proximity. However, certain mixtures of organizationally "close" and "distant" participants in an interaction result in more effort needed to communicate. Also, interactions tend to be more effort-intensive when they occur in a meeting and when more people are involved. These results provide a better understanding of how organizational structure helps or hinders communication in software development.

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[1] 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 one another 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 that influence the amount of effort software developers spend in communication activities. This study is a first step toward providing management support for control of communication effort.

The empirical study described here aims to identify the organizational characteristics that 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 independent variables are organizational distance, physical distance, and familiarity. These three variables 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 mea-

0018-8670/97/$5.00 © 1997 IBM

sures are based are, respectively, official relationships, physical proximity, and past and present working relationships.

The study combines quantitative and qualitative research methods. It is based on data collected from the observation of 10 (mostly design and code) review meetings and from interviews with the review participants. This study is exploratory in nature, in that no formal hypotheses were tested. On the contrary, the goal of the study is the generation of theory, and one of its contributions is a set of proposed hypotheses to guide further study.

Several interesting findings were discovered concerning organizational relationships among review participants and the amount of effort involved in different types of interactions among them. Not surprisingly, people took less time to communicate when they were familiar with one another and when they worked in close physical proximity to one another. However, interactions tended to take longer when they involved a group of organizationally "close" participants, plus a few "outsiders." These types of interactions took longer than when the interaction was between all organizationally close participants, or all organizationally distant participants. Interactions also tended to require more effort when they occurred during a meeting, which implies that participants did not cut short their discussions in order to make the meeting shorter. More obviously, interactions involving more people generally took longer.

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,[2,3] to organizational growth,[4] to the study of technological organizations,[5] to business process re-engineering.[6,7] Most of this work takes a global, long-range perspective. The common proposition is that organizational structure evolves over time to facilitate communication and that, in fact, facilitating communication is the main purpose of organizational structure. Organization theory also states the benefits of organizational and physical proximity of communicators,[5,8] as well as the importance of various types of personal relationships (such as familiarity as we have defined it) among communicators.[5] However, organization and communication have not been explored in detail in software development organizations. Several studies have provided evidence of the relevance of both organizational structure (along with other "nontechnical" factors) and communication in software development. In particular, at least one study[1,9] has shown the drawbacks of organizational and physical distance and discovered the importance of "shared internal representations," which has led to our particular definition of familiarity.

Study setting

This study took place at the IBM Software Solutions Laboratory in Toronto, Canada. The development

Organization and communication have not been explored in detail in software development organizations.

project studied was DB2* (DATABASE 2*), an IBM commercial database system with several versions for different platforms. During the month of June 1994, data were collected from 10 (mostly design and code) reviews in Toronto. Observations of the review meetings were followed up with interviews with review participants in November 1994 and 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 detailed description of the DB2 development environment, written for the purposes of this study, was used extensively throughout the investigation. 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.

The description had three major parts, corresponding to the three aspects of the environment most relevant to this study: the review process, the organizational structure, and information flow. These parts are described in the following subsections.

Process. The official purpose of the DB2 review process, as stated in the review process documentation, is twofold. One objective is to discover defects early in the software development process. The other is to discover defects that the author of the code is likely to overlook, by exposing the code to a wider group of people. Another stated but unofficial benefit of reviews is the sharing of knowledge and team building.

The DB2 review process consists of steps for planning, preparing, meeting, rework, and follow-up. Reviewers (anywhere from 2 to 13) are picked during the planning step; they then read and review the material individually during preparation. The meeting is generally spent raising and discussing the defects that the reviewers discover, as well as discussing other issues about the reviewed material. During the rework step, the original author is responsible for correcting the defects found. One reviewer, designated the chief reviewer, is responsible for ensuring that the corrections are made satisfactorily during the follow-up.

Most developers interviewed expressed a strong belief in and commitment to reviews. They seemed to take the process seriously in general, although they viewed some process steps (mainly administrative) as unimportant. However, they felt that properly preparing for each review was vitally important and expressed annoyance when a reviewer appeared to be unprepared, or when the material to be reviewed was not really ready to be reviewed. Also important was choosing the "right" mix of reviewers. The review meeting itself was not always seen as a crucial part of the process, and developers had differing opinions about the best way to conduct the meeting. Flexibility in the process was seen as valuable.

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, a review is conducted for each major artifact (requirements, design, code, and test cases). In this study, we observed and measured one requirements review, five design reviews, three code reviews, and one test case review.

Organization. The formal DB2 organization has a basic hierarchical structure. First-line managers manage small teams of developers who may be responsible for maintaining specific collections of software components, facilitating releases of DB2 products, or providing support services, such as system test, to all the other teams. Some teams of developers are further divided into groups headed by a task leader. One second-line manager is responsible for all DB2 development.

Other organizational relationships included in the description are those corresponding to work patterns and physical locations. In particular, groups are defined that include developers who work together on a regular basis and are familiar with one another's work. Also, people are grouped who share offices, corridors, buildings, and sites.

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

We identified 13 different types of interactions that take place during the review process. One interaction type, for example, called *defects*, refers to the raising and discussing of defects during the review meeting. Another, quite different, example is the *schedule_meeting* interaction type, in which the author of the material to be reviewed must negotiate with all the reviewers to find a suitable meeting time. Not all 13 interaction types occurred during every review. Also, not all interaction types are relevant to the results presented in this paper, so only a few of them are described in detail (in the section on results).

Study design

This empirical study examines the role of organizational structure in process communication among software developers. Our research design combines qualitative and quantitative methods. Qualitative methods were used to collect data which were then quantified, or coded, into variables that were analyzed using quantitative methods.

The unit of analysis in this study is the *interaction*, 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, as opposed to unplanned interactions that do not correspond directly to some step of a defined process.

Although such informal communication has been shown to be more valuable,[10] there is still a need for focused studies of process communication because it can be planned for and controlled if we know the factors that can be manipulated to make it more efficient.

The data collection procedures used in this study are participant observation[11] and structured interviews.[12] The data gathered from these different sources overlap, thus providing a way of triangulating,[12] or cross-checking the accuracy of the data. After the data were

The choice of dependent variable is motivated by the goals of the study.

collected, they were coded and transformed into a set of quantitative variables, such as the number of people present at the observed meeting, the lengths of different types of interactions that took place, and the type of communication medium used.

Data collection. The data for this study were collected during an initial visit to IBM in Toronto in June 1994 and two follow-up visits in November 1994 and April 1995, and with several e-mail communications between the visits. The bulk of the data came from direct observation of 10 reviews of DB2 line items in June 1994. Administrative information pertaining to each review (e.g., participants, date, time, material reviewed, preparation time) was collected from several sources, thus providing opportunities for triangulating the data. These sources included review announcements, IBM data collection forms, and observed statements during the review meeting. During every observed review, each separate discussion was timed. The beginning and ending times, the participants, and the type of each discussion were recorded.

Interviews with review participants were also conducted to elicit information about other parts of the review process that were not observed, to clarify discussions that took place during the meeting, and to triangulate data that had already been collected. The initial interviews were conducted within a few days of each DB2 review. Other interviews took place later in November 1994 and April 1995. For each review, an author and several reviewers were interviewed.

Variables. The variables chosen for analysis fall into three categories. First is the dependent variable, *communication effort.* Second, a set of independent variables represent organizational structure, defined as the network of relationships between review participants. Several different measures have been chosen that are based on different types of organizational relationships. Finally, a large set of intervening variables threaten to confound the results if not taken into account.

The dependent variable, labeled *CE* (for communication effort), is the amount of effort, in person-minutes, expended to complete an interaction. How *CE* is determined for each interaction depends on the interaction type. For example, for interactions that take place during the review meeting (e.g., *defects* type interactions), the information recorded during observations about individual discussions is used to calculate *CE*, along with information about preparation time. Calculation of *CE* for some other types of interactions depends on interview data.

The choice of dependent variable is motivated by the goals of the study. One objective is to consider aspects of communication that might be affected by organizational factors, i.e., effort. Another is to provide useful information that can be used for managing and planning, hence the emphasis on process communication as opposed to informal, unplanned communication.

There are four organizational structure variables. The first two, *XOD* and *MOD*, are both based on *organizational distance*, which quantifies the degree of management structure between two members of the organization. It is calculated as the length of the shortest path between the two in a graph representing the reporting structure of the organization (the "org chart"). *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 many of the participants are organizationally distant. At IBM, the management structure tended to change frequently, in part because there was an effort to make it reflect actual work patterns in the organization, which tended to shift over time. Therefore, organizational distance provides a more accurate picture of working relationships in this setting than it would in an organization in which the reporting structure is more stable.

The two other organizational variables are *familiarity* (*Fam*) and *physical distance* (*Phys*). 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. This factor was found to be relevant in the study by Krasner et al.[9] of large software development projects. The familiarity measure also attempts to capture the important informal networks that do not make up part of the management structure (measured by organizational distance). It is calculated first as the percentage of pairs of interaction participants who have indicated (in interviews) that they are familiar with each other's work, based on past or present working relationships. It is then coded into a four-level ordinal variable. So a high value for familiarity means that most of the interaction participants have worked together in the past or currently work together on a regular basis.

Physical distance reflects the number of physical boundaries (walls, buildings, cities) between the interaction participants, which was mentioned as important by numerous people in the study setting. The variable is ordinal and in ascending order. It has four levels corresponding to, respectively, interaction participants who share an office, who have offices on the same corridor, who have offices in the same city, and who work in different geographical locations.

Although the set of intervening variables in this study is large, it is still probably inadequate to capture the full richness of the study context. A number of factors outside the scope of this study, such as culture or personality types, may have significantly influenced communication effort. However, the chosen intervening variables represent a reasonable compromise between practical concerns (availability of data, measurability, researcher expertise, etc.) and an effort to eliminate confounding factors. A few of the intervening variables most relevant to the results presented in this paper are listed below:

- *N* is the size of the set of interaction participants, or the number of people who expend effort in an interaction.
- *Mr* is the communication medium used to request the information shared in an interaction; it can be N/A, verbal, or e-mail.
- *Mp* is the communication medium used to prepare the information to be shared; it can be a paper form, verbal, a short written message, a structured document, or an unstructured document.
- *Mt* is the communication medium used to transfer the information between participants; it can be a face-to-face meeting, a conference call, a video conference, an electronic transfer (e-mail, FTP, etc.), on paper, or a normal telephone call.
- *Size* is the amount of information shared in an interaction; it is measured on an ordinal scale that folds together measures in pages and lines of code.
- *Type* is the kind of interaction, e.g., *defects*.

Data analysis. Our analysis method involved partitioning the data based on the values of one or more variables, then analyzing those subsets. The subsets of interactions that we analyzed are:

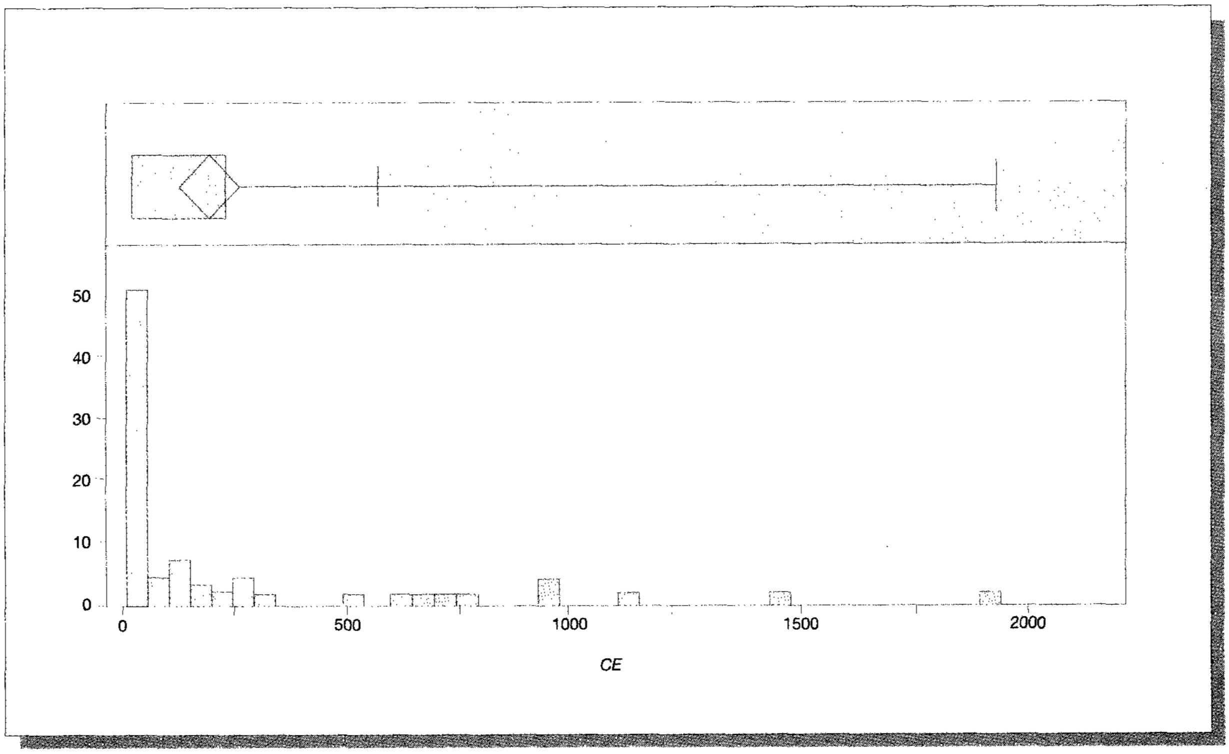
- The entire set of interactions
- High-effort interactions ($CE > 500$)
- Technical interactions that take place during the review meeting (interaction types—*questions*, *defects*, and *discussion*)
- Partitioned by the 13 interaction types
- Partitioned by combinations of the organizational variables (e.g., low physical distance and high *MOD*)

For each of these subsets, histograms (to show distributions of one variable) and scatterplots (to show relationships between *CE* and the organizational variables) were generated. To test relationships, Spearman correlation coefficients were calculated. Another two-variable relationship that we explored with scatterplots is the relationship between *CE* and the number of participants (*N*). We also ran ANOVAs (analysis of variance tests) on some combinations of variables for some subsets, but there were 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.

Results

Space does not permit a presentation of the results of all the analyses described in the previous section. Instead, we present three of the most interesting sets of findings in the form of hypotheses that we have

Figure 1 **Distribution of communication effort (in person-minutes) over all 100 interactions**



generated and the evidence supporting them. Before that, however, is a brief characterization of the scope of the data, in particular the distributions of the dependent and independent variables.

Data characterization. First, as can be seen in Figure 1, the distribution of the dependent variable, communication effort, is highly skewed toward 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 percentiles, and the 90th percentile and upper extreme are shown as vertical lines. The diamond indicates the mean of the data. Ninety percent 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 three person-minutes. The median was 38 and the mean was about 190.

Table 1 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 percent of the interactions had a median organizational distance (*MOD*) of two or less, and more than three-fourths had a maximum organizational distance (*XOD*) of four or higher. If we look at *MOD* and *XOD* together, as in Table 2, we see that most of the data fall into three categories:

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

Familiarity and physical distance. In this subsection, we present two hypotheses.

Hypothesis: Interactions tend to require more effort when the participants are not previously familiar with one another's work.

Figure 2 Distribution of familiarity over all 100 interactions

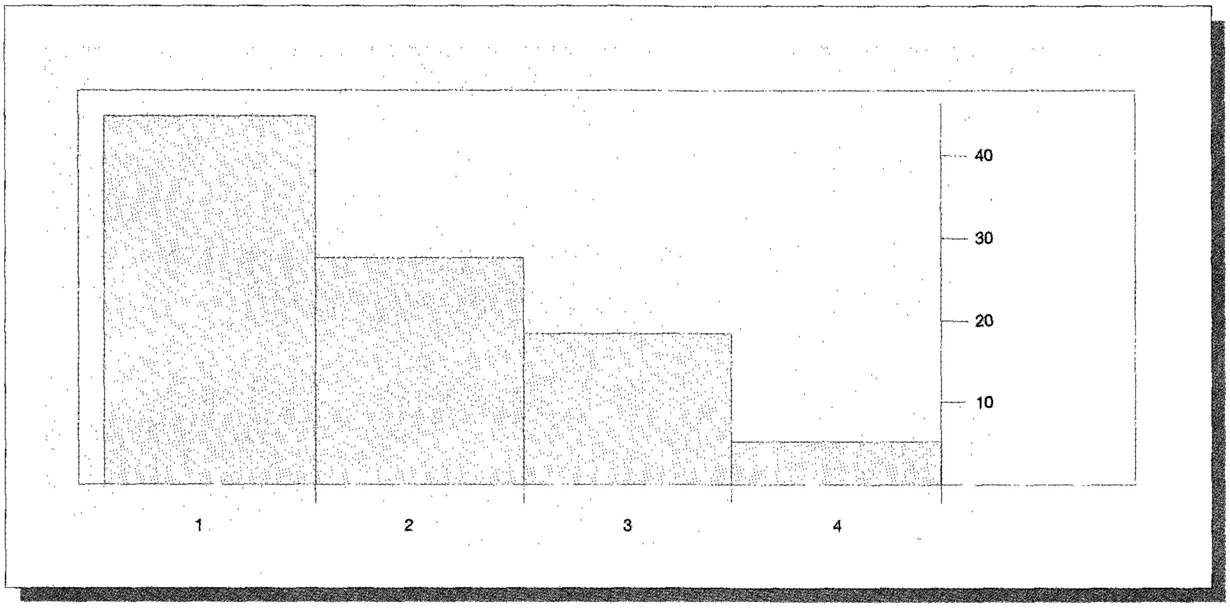


Table 1 Frequency table for median (*MOD*) and maximum (*XOD*) organizational distance

Level	*MOD*		*XOD*	
	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

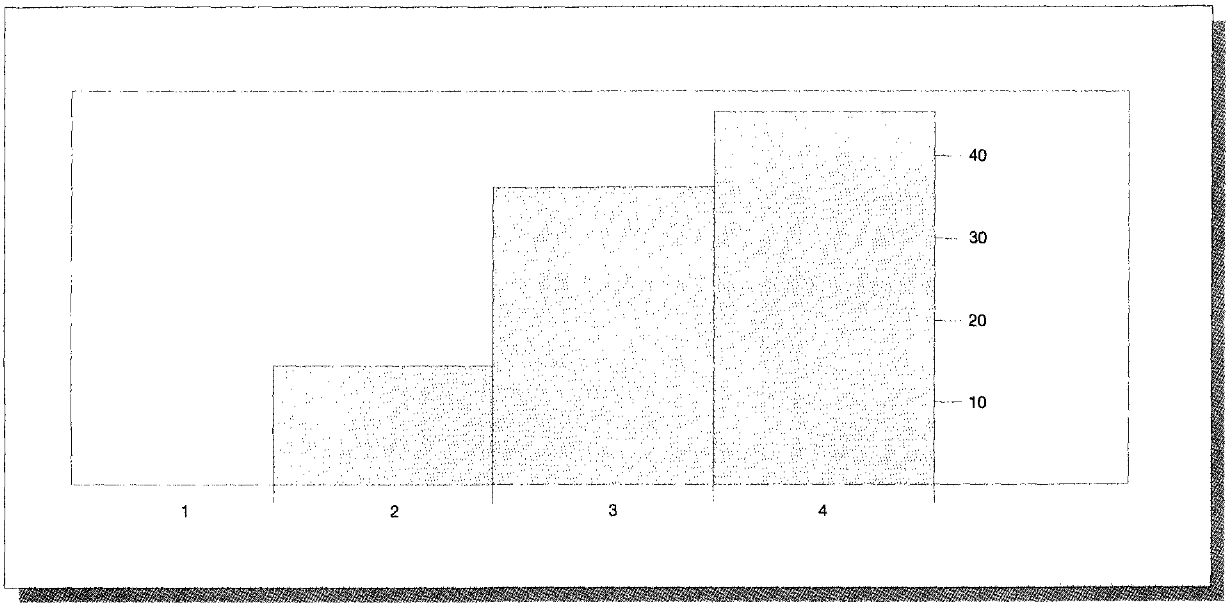
Table 2 Frequency of values for median and maximum organizational distance

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

Hypothesis: Interactions tend to require more effort when the participants work in physically distant locations.

Two of the organizational structure variables, familiarity and physical distance, exhibit straightforward relationships with communication effort. Figures 2 and 3 show their distributions. In Figure 4, they are both plotted against communication effort. A box plot is also shown for each level of each independent variable. The top and bottom boundaries of the boxes indicate the 75th and 25th percentiles. The median and the 90th and 10th percentiles are also shown as short horizontal lines (the median and 10th percentiles are not really visible on most boxes). The width of each box (and of the partitions on the horizontal axis) reflects the number of data points in that level. It appears from Figure 4 that high effort is associated with low familiarity and with high physical distance (the latter observation being the strongest). That is, interactions tend to require more effort when the participants are not previously familiar with one another's work. This observation is consistent with Krasner's findings[9] about "common internal representations." As for physical distance, interactions tend to require more effort when the participants work in physically distant locations. Curtis[1] and Allen[5] have had similar findings. However, it must be noted that most interactions have low familiarity and high physical distance, as shown in Figures 2 and 3.

Figure 3 Distribution of physical distance over all 100 interactions



The Spearman correlation coefficients, which reflect the strength of the relationships between each independent variable and the dependent variable, are shown in Table 3. Physical distance has the highest coefficient, which implies that it has a stronger direct effect on communication effort than any of the other variables.

Table 3 Spearman rho (ρ) coefficients comparing each independent variable to the dependent variable, *CE*

MOD	*XOD*	*Fam*	*Phys*
0.09	0.4	0.14	0.5

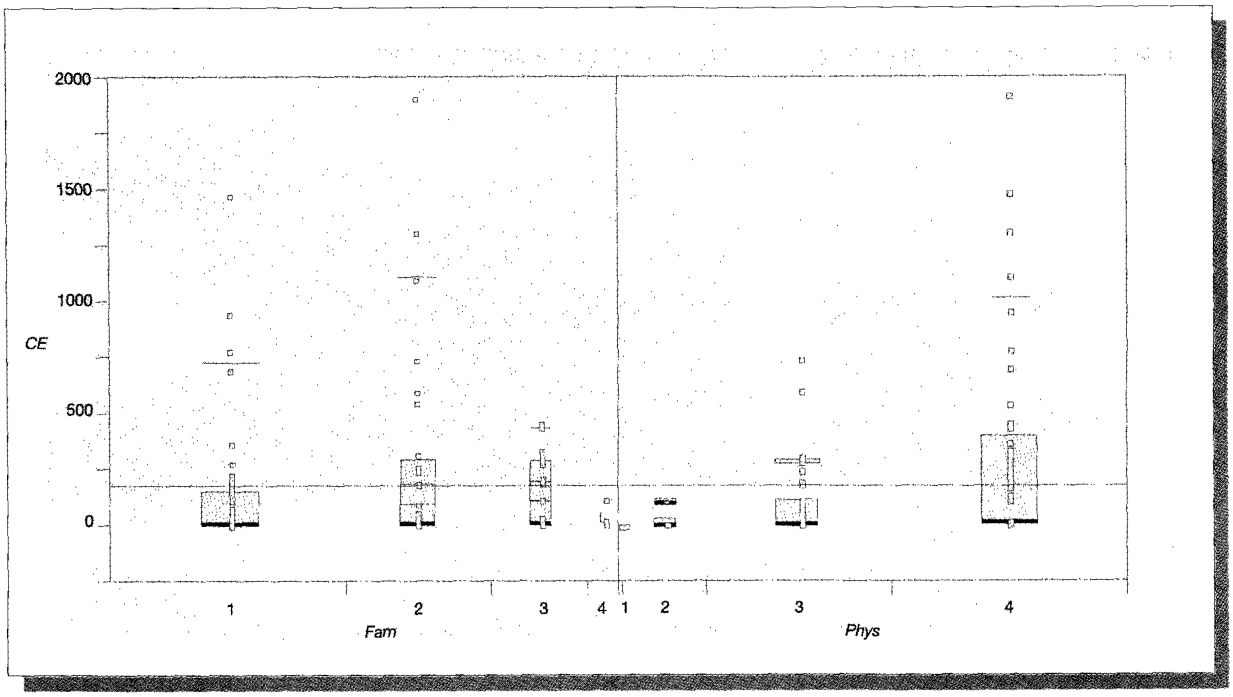
Organizational distance. We now present a hypothesis on organizational distance.

Hypothesis: More effort is required when the set of participants includes mostly organizationally close members, but has a few organizationally distant members.

The results pertaining to organizational distance are both more complex and more interesting. Figure 5 shows two scatter plots, each with communication effort on the vertical axis, and one of the two versions of the organizational distance variable on the horizontal axis. Box plots for each level of each independent variable also show the 10th, 25th, 75th, and 90th percentiles, as well as the median.

From Figure 5, 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 category is the second described in the subsection on familiarity and distance, 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 significance of this result by calculating the Mann-Whitney U statistic. This is a nonparametric 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 that did not. The test yielded a significant value, even

Figure 4 Communication effort plotted against familiarity (*Fam*) and physical distance (*Phys*)



at the $p < .01$ significance level. This result contrasts with Curtis,[1] who hypothesized that the relationship between organizational distance and communication ease is more straightforward.

If we examine in more detail the subset of interactions that were effort-intensive, we find further evidence supporting this hypothesis. The 11 highest-effort interactions all required a communication effort greater than 500 person-minutes. The distributions of organizational distance in this subset are shown in Table 4. None of the high-effort interactions had a *MOD* more than two, and none had a *XOD* less than four. In fact, all of the interactions in this high-effort subset belong to the second category (low *MOD*/high *XOD*) described above.

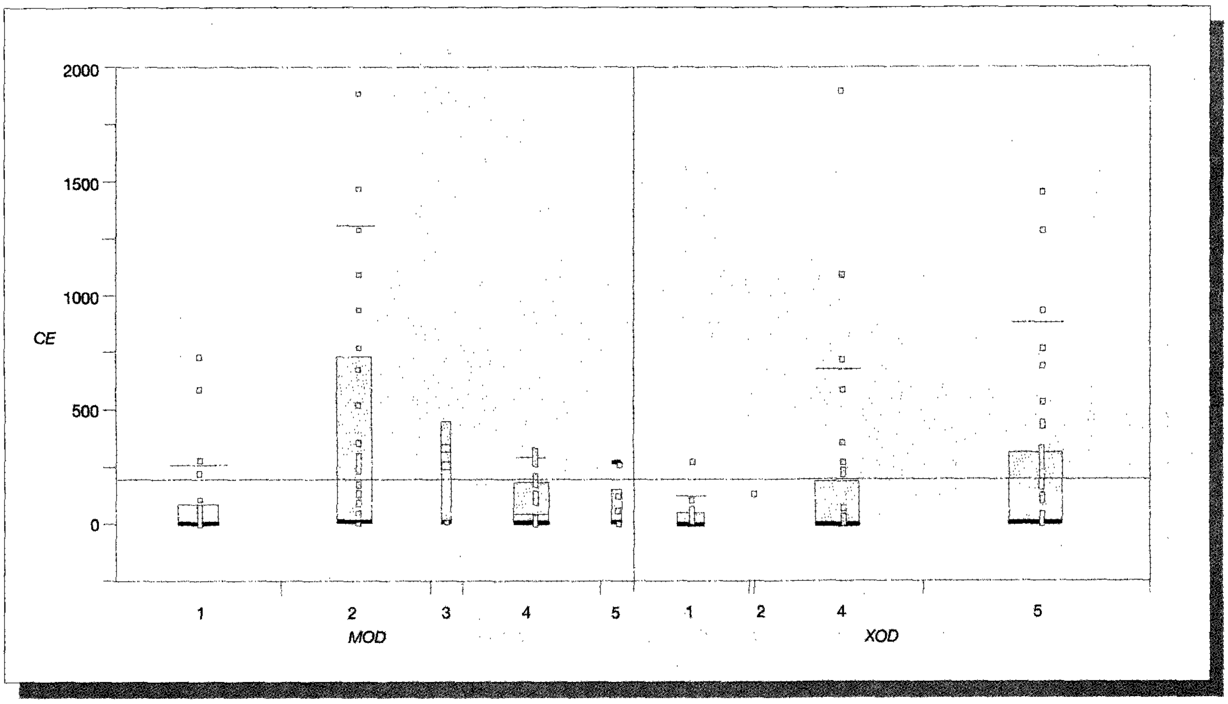
Three of the types of interactions that were most effort-intensive were the *defects*, *questions*, and *discussion* interaction types. Most of the interactions of type defects involved a set of participants that fell into the second category (low *MOD*/high *XOD*), including all of those with *CE* above the mean. Recall that defects interactions are those in which review participants raise and discuss defects during the review meeting. The questions interaction refers to questions raised and discussed during the review meeting. Again, the highest-effort interactions of this type fall into the second category of participant sets (low *MOD*/high *XOD*). The discussion interactions (which include other types of technical discussion during the review meeting) 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 distance as mentioned above for the questions and defects interactions.

Meeting interactions. Two hypotheses are presented for meetings.

Hypothesis: Interactions that take place during a meeting (a verbal request and an unprepared reply) tend to require more effort than other interactions, especially when they involve communication technology.

Hypothesis: Having more participants tends to make interactions more effort-intensive, even when the effort is normalized by the number of participants.

Figure 5 Communication effort plotted against median organizational distance (*MOD*) and maximum organizational distance (*XOD*)



Nearly all of the high-effort interactions involved a verbal request for information (Mr = verbal) and no written preparation of the information (Mp = verbal), and were executed using some sort of communications technology (Mt = videoconference or conference call). These patterns in the use of communication media, shown in Figure 6, differ dramatically from the patterns seen in the data as a whole. Interactions that 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. This result implies, however, that the meeting participants did not, or could not, keep the meeting from running long by cutting short their discussions.

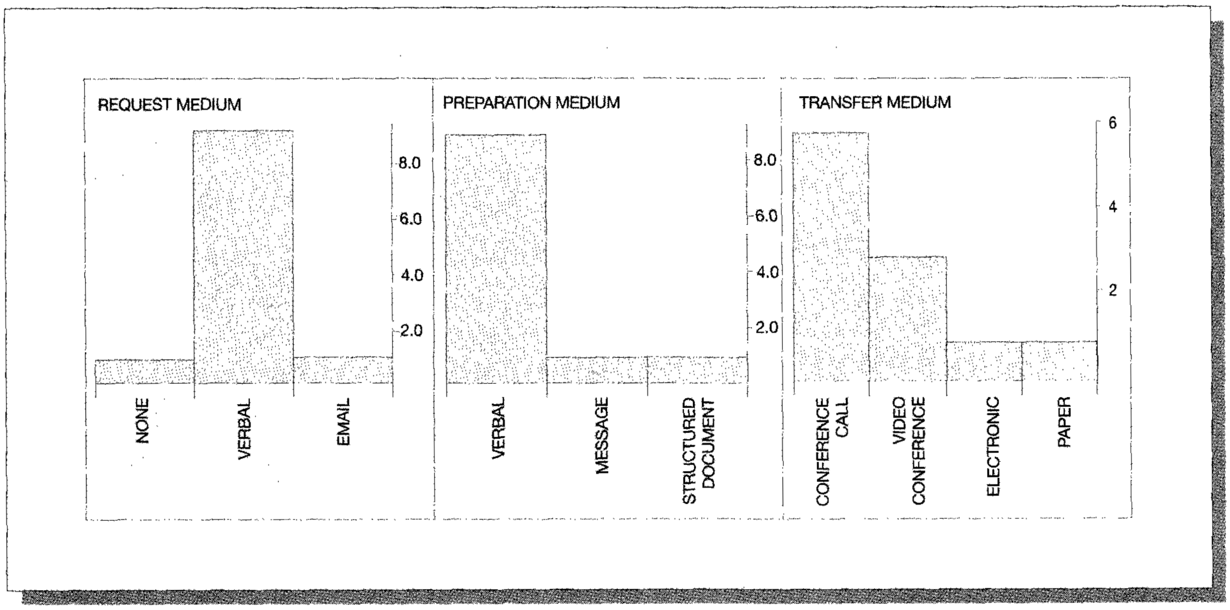
The defects, questions, and discussion interactions constitute all of the technical communication that takes place during a review meeting. The effort recorded for these interactions includes the effort required to prepare for, carry out, and digest this technical information. Since these interactions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

Table 4 Frequency of values for median and maximum organizational distance for the 11 highest-effort interactions

MOD	XOD	
	4	5
1	2	0
2	2	7

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

Figure 6 The different communication media used in the 11 highest-effort interactions



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 an 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 that included the 11 interactions we analyzed before as the highest effort). The median number of participants in this subset is 8, lower 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.

This result should not be surprising, given that it is consistent with the theoretically quadratic growth of the number of communication channels among n people. That is, there are $n\ (n-1)/2$ pairs among n interaction participants, so, intuitively, the effort to communicate in a group of n should also grow faster than n.

Defects interactions included anywhere from 4 to 15 participants, with a median of 10 participants. All of the defects interactions with (normalized or unnormalized) CE above the mean had seven or more participants. The same was true for *questions* interactions.

Lessons learned

One of the goals of this study was to serve as a pilot for a larger study. Although small, this pilot study was valuable in clarifying a number of issues related to how this subject is best studied. The limitations discussed below have been remedied in the design of the larger study, resulting in a more useful and relevant study that was conducted more recently at the NASA (National Aeronautics and Space Administration) Goddard Space Flight Center under the auspices of the Software Engineering Laboratory.[13] The main differences between the larger 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 of this study, as described below. The main goal of the larger study, as in the pilot study, is to learn how organizational structure characteristics affect the amount of effort expended on communication.

The major limitations of this pilot study are 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 the statistical significance of many of the findings or to generalize the results in any way. A better study design would include more data points, although reducing the number of variables would reduce the need for a very large data set.

The three-part description built at the beginning of the study (described in the second section) 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 somewhat limited the analyses that could be done easily. Some automatic support for managing this part of the description (or even handling the data through an electronic form of the description), as well as a better notation, was needed. In particular, we have used Yu's Actor-Dependency models[14] as a notation and mechanism in subsequent studies for organizing information and data, and for automating some of the data analysis.

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 aspect 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. For the larger study, we have chosen the review itself as the unit of analysis. In addition, more care was taken to model the process as it was enacted, not necessarily as it was defined.

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 that has been used in the larger study.

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 during a period of 10 months. This length of time led to memory, personnel turnover, and discontinuity problems. A single interview, done as shortly after the review as possible, is preferred, which we were able to do in the larger study.

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, which sometimes made it difficult to determine what type of discussions were taking place during observations. In addition, no reliability techniques were employed, such as audio- or videotaping the reviews, or having a second observer. Using such techniques would have ensured better accuracy of the data. The latter method was used in the larger study to assess the subjectivity and accuracy of the information recorded during the observations. Also, with regard to data accuracy, some variables had no triangulation[12] 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. In the larger study, more data sources were used, but still not every data element could be triangulated.

During observations and interviews, some field notes were taken in addition to the information on the interview forms and observation checklists. However, these data were not extensive or reliable enough to be used as part of the data analysis. If more faithful notes had been kept, such 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 a little the way in which the interviews were structured. Field notes were used extensively in the larger study.

A better study design, then, which results from the experience of this pilot study, would still be based on qualitative data from interviews and observations. However, these data would be better managed through an automated and consistent model, and better exploited through more extensive use of field notes and qualitative analysis. Also, the accuracy of the data would be greater because of reliability techniques (a second observer or recording of some kind), more timely interviews, and better triangulation. The better study would also be larger and would begin with a period of prior ethnography in order to improve the design of the variables and more accurately model the setting.

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.

The goal of this study was the generation of theory, and so the findings are best presented as proposed hypotheses. The study results 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 the relevant factors affecting communication in software development:

- Interactions tend to require more effort when the participants are not previously familiar with one another's work.
- Interactions tend to require more effort when the participants work in physically distant locations.
- More effort is required when the set of participants includes mostly organizationally close members, but with a few organizationally distant members.
- Interactions that take place during a meeting (a verbal request and an unprepared reply) tend to require more effort than other interactions, especially when they involve communication technology.
- Having more participants tends to make interactions more effort-intensive, even when the effort is normalized by the number of participants.

The independent variables in this study can all be said to measure different aspects of "closeness," and one might expect the findings to straightforwardly confirm what is intuitive—that "closer" people communicate more easily. The first two statements above, in fact, do rather confirm the obvious (which is still a valuable use of empirical evidence). However, the third hypothesis reflects the fact that there is some complexity to the issue of how organizational structure affects communication between developers. The last two hypotheses question the value of meetings, especially large meetings, which could be further investigated as a separate issue.

This study is not sufficient to solve the problem of managing information flow in a software development organization and process, but it is a first step. The findings of this study could be used to help in planning, for example, by pointing out characteristics that increase communication costs in reviews. They might also help in diagnosing communication problems as they arise.

But the most important contribution of this work is that it enables the further exploration of this area, and provides methods and proposed hypotheses with which to begin. This study represents a very small first step in building the experience necessary to effectively manage information flow in software development organizations. It also exemplifies a scientific approach that is applicable to the study of numerous issues in software engineering and that would help the field to mature toward a true engineering discipline. The next step for the authors is the larger empirical study described briefly in the last section. 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. In addition, this study does not address the issue of communication *quality*, only *quantity*. One cannot assume that the two are equivalent. Finally, more work is necessary in the area of actually applying this new knowledge to the improvement of software development projects and the mechanisms needed to achieve such improvement.

Acknowledgment

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

Cited references

1. B. Curtis, H. Krasner, and N. Iscoe, "A Field Study of the Software Design Process for Large Systems," *Communications of the ACM* **31**, No. 11, 1268–1287 (November 1988).
2. J. R. Galbraith, *Organization Design*, Addison-Wesley Publishing Co., Reading, MA (1977).
3. J. G. March and H. A. Simon, *Organizations*, John Wiley & Sons, Inc., New York (1958).
4. A. L. Stinchcombe, *Information and Organizations*, University of California Press, Berkeley, CA (1990).
5. T. J. Allen, *Managing the Flow of Technology*, The MIT Press, Cambridge, MA (1985).
6. T. H. Davenport and J. E. Short, "The New Industrial Engineering: Information Technology and Business Process Redesign," *Sloan Management Review*, 11–27 (Summer 1990).
7. M. Hammer, "Reengineering Work: Don't Automate, Obliterate," *Harvard Business Review*, 104–112 (July 1990).
8. H. Mintzberg, *The Structuring of Organizations*, Prentice-Hall, Inc., Englewood Cliffs, NJ (1979).
9. H. Krasner, B. Curtis, and N. Iscoe, "Communication Breakdowns and Boundary Spanning Activities on Large Programming Projects," *Empirical Studies of Programmers*, Second Workshop, Chapter 4, G. Olsen, S. Sheppard, and E. Soloway, Editors, Ablex Publishing, Norwood, NJ (1987), pp. 47–64.
10. R. E. Kraut and L. A. Streeter, "Coordination in Software Development," *Communications of the ACM* **38**, No. 3, 69–81 (March 1995).
11. S. J. Taylor and R. Bogdan, *Introduction to Qualitative Research Methods*, John Wiley & Sons, Inc., New York (1984).
12. Y. S. Lincoln and E. G. Guba, *Naturalistic Inquiry*, Sage, Beverly Hills, CA (1985).
13. C. B. Seaman and V. R. Basili, "An Empirical Study of Communication in Code Inspections," *Proceedings of the 1997 International Conference on Software Engineering*, Boston, MA (May 17–24, 1997).
14. E. S. Yu and J. Mylopoulos, "Understanding 'Why' in Software Process Modeling, Analysis, and Design," *Proceedings of the 16th International Conference on Software Engineering*, Sorrento, Italy (1994), pp. 159–168.

Accepted for publication June 3, 1997.

Carolyn B. Seaman *Computer Science Department, A. V. Williams Building, University of Maryland, College Park, Maryland 20742 (electronic mail: cseaman@cs.umd.edu).* Dr. Seaman is a faculty researcher in the Experimental Software Engineering Group at the University of Maryland. She has been conducting empirical research in software engineering for over five years. While working on her doctorate, she received a fellowship from IBM Canada's Centre for Advanced Studies. Previously, she had worked as a software developer in a number of different environments. She received her Ph.D. from the University of Maryland in 1996, her M.S. from the Georgia Institute of Technology in 1987, and her B.A. from the College of Wooster (Ohio) in 1986.

Victor R. Basili *Institute for Advanced Computer Studies, Computer Science Department, A. V. Williams Building, University of Maryland, College Park, Maryland 20742 (electronic mail: basili@cs.umd.edu).* Dr. Basili is a professor in the Institute for Advanced Computer Studies and the Computer Science Department, where he served as chairman for six years. He works on the development of quantitative approaches for software management, engineering, and quality assurance, using models and metrics for improving the software development process and product. He is one of the founders and principals in the Software Engineering Laboratory, a joint venture between the NASA Goddard Space Flight Center, the University of Maryland, and Computer Sciences Corporation, established in 1976, and recipient of the first Process Improvement Achievement Award by the IEEE and the SEI in 1994 and a NASA Group Achievement Award in 1996. He is also a recipient of a NASA/GSFC Productivity Improvement and Quality Enhancement Award. He received the 1997 Award for Outstanding Achievement in Mathematics and Computer Science given by the Washington Academy of Sciences. Dr. Basili holds a B.S. from Fordham College, an M.S. from Syracuse University, and a Ph.D. from the University of Texas at Austin. He has authored over 130 journal and refereed conference papers. He served as Editor-in-Chief of the *IEEE Transactions on Software Engineering*, Program Chair and General Chair of the 6th and 15th International Conference on Software Engineering, respectively, and for several other conferences. He is coeditor-and-chief of the *International Journal of Empirical Software Engineering*. He serves on the editorial board of the *Journal of Systems and Software* and is an IEEE and ACM Fellow.

Reprint Order No. G321-5658.