An Empirical Study of Communication in Code Inspections

Carolyn B. Seaman
Institute for Advanced Computer Studies
Computer Science Department
University of Maryland
College Park, MD, USA
(301) 405-2721
cseaman@cs.umd.edu

Victor R. Basili
Institute for Advanced Computer Studies
Computer Science Department
University of Maryland
College Park, MD, USA
(301) 405-5497
basili@cs.umd.edu

ABSTRACT

This paper describes an empirical study which addresses the issue of communication among members of a software development organization. In particular, data was collected concerning code inspections in one software development project. The question of interest is whether or not organizational structure (the network of relationships between developers) has an effect on the amount of effort expended on communication between developers. Both quantitative and qualitative methods were used, including participant observation, structured interviews, generation of hypotheses from field notes, some simple statistical tests of relationships, and interpretation of results with qualitative anecdotes. The study results show that past and present working relationships between inspection participants affect the amount of meeting time spent in different types of discussion, thus affecting the overall meeting length. Reporting relationships and physical proximity also have an effect, as well as the point in the project that an inspection occurs. All but the last of these factors are organizational structure relationships. The contribution of the study is a set of well-supported hypotheses for further investigation.

Keywords

empirical study, communication, process, organizational structure, inspections

INTRODUCTION

Many factors which impact the success of software development projects still defy our efforts to control, predict, manipulate, or even identify them. One factor that has been identified [3] 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 effectively) [12]. It is also clear that efficient information flow is affected by the relationship between development processes and the organizational structure in which they are executed. 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. These relationships between general concepts are pictured in Figure 1.

The study described in this paper addresses the productivity aspects of communication. In particular, it empirically studies how process communication effort (the effort associated with the communication required by a development process) is influenced by the organizational structure (the network of relationships between developers) of the development project. In this paper, we examine the organizational structure of one particular project, the code inspection process used, and the time and effort associated with inspection meetings. We found that organizational attributes are significantly related to the amount of time inspection participants spend in different types of discussions. The aim of this study is not to test or validate hypotheses about relationships between these variables, but to explore what relationships might exist and try to explain those relationships. Our contribution, then, is a set of *proposed* hypotheses, along with an argument, in the form of supporting evidence, for their further examination.

Although the importance of efficient communication, and its relationship to organizational structure, is well supported in the organization theory literature [11, 5], it has not been adequately addressed for software development organizations. Communication has been identified as an important factor in how developers spend their time [12], and some organizational characteristics which affect its efficiency have been suggested [3, 8]. Some, but surprisingly little, of the "process" work in software engineering has dealt with information flow or organizational structure [1, 2, 13]. It has been postulated that informal communication is usually more

ICSE 97 Boston MA USA

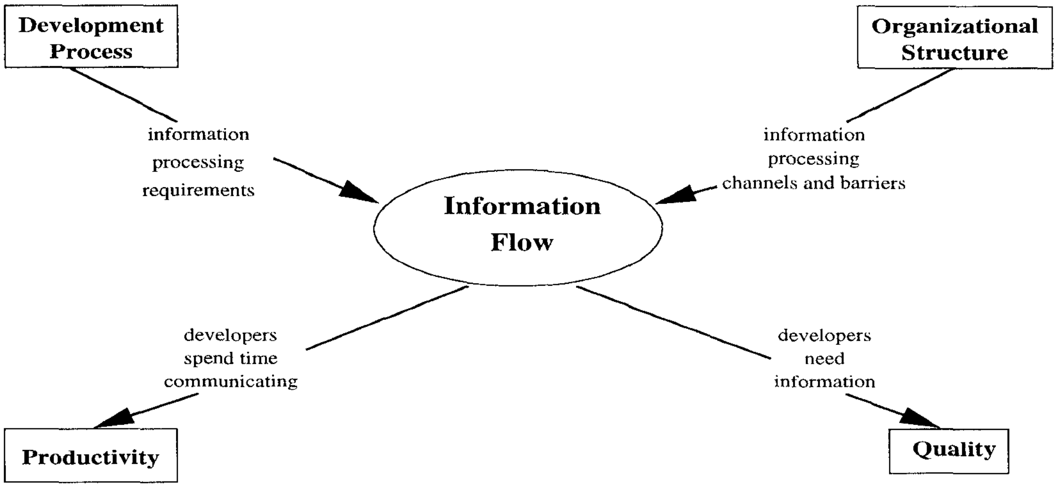

Figure 1: Relationships between concepts relevant to this work.

valuable than formal, interpersonal communication [9] (which includes what we have termed process communication). However, there is still a need for focused studies of the latter because, unlike informal communication, formal communication can be planned for and controlled, if we know the factors which can be manipulated to make it more efficient. Studies of human communication must, by definition, be empirical studies because they deal with non-analytical entities (i.e., people) which have few universally applicable laws or theories governing their behavior. Concepts such as communication, process, and organization must be studied where they occur, in real software development projects.

The study combines quantitative and qualitative research methods. Qualitative data is data represented as words and pictures, not numbers [6]. Qualitative methods are especially useful for generating, rather than testing, hypotheses. 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 such as this one, but only where well-defined quantitative variables are being studied. We have combined these paradigms in order to flexibly explore an area with little previous work, as well as to provide compelling evidence to support the hypotheses we present.

STUDY SETTING

The project used for this study involved the development of mission planning software for NASA/Goddard Space Flight Center's Flight Dynamics Division (FDD). Much of the development was contracted to Computer Sciences Corporation (CSC). About 20 technical leads and developers (most from CSC) participated in the inspection process during the course of the study, although more participated in the project. The project began in early 1995, and the first release was scheduled for the summer of 1996 (as of this writing, it has not yet been delivered).

The two aspects of the project which are of interest are the development processes used (in particular the communication required by those processes), and the organizational relationships between the process participants.

This study focuses on the project's code inspection process. We relied initially on a written document, the *Code Inspection Procedure*, which defined the tailored inspection process for this project, including the relevant steps and roles. Throughout the study, however, we updated our understanding of the inspection process through observation and interviews. Inspections were conducted after unit test, before submitting the code to configuration control. Both code and unit test products (test plan and results) were inspected. It should be noted that some of the code inspected was produced by a code generator, which was used to write skeletons for all the classes developed, and some of the supporting code. The inspection meeting (the unit of analysis for this study) was one step in the inspection process. Inspection meeting participants included the "author", who had implemented and unit tested the C++ classes being inspected, the "moderator", a "code inspector", and a "test inspector". In some cases, more developers were assigned to inspect the code or test. All the observed inspections occurred at CSC, and involved mostly CSC personnel. The objective of the inspection meeting was to record defects which had been found by the inspectors during their preparation.

We have defined organizational structure as a network of organizational relationships. These relationships include management, or reporting, relationships and physical proximity. Information about these types of relationships came mainly from organizational documents, and was validated through interviews. Other organizational relationships we studied were past and present working relationships, which existed between many of the CSC and FDD personnel.

The organization and process information described above was modeled using a formalism called *Actor-Dependency models* (AD models) [15]. This model also included the actual data collected, and allowed us to automate some of the data analysis.

The process used to produce the AD model is known as *prior ethnography* [10], the practice of taking some time before data collection begins to become familiar with the study setting. In November and early December of 1995, the researcher attended team meetings, conducted open-ended interviews with several developers and managers, observed several inspection meetings (without recording data), and was introduced to all project participants. The goal was to become familiar with the setting, produce the AD model, and choose the relevant dependent and independent variables.

RESEARCH METHODS

The data and methods employed in this study are both quantitative and qualitative. The data collection phases were largely qualitative. The qualitative part of the data analysis began about halfway through data collection, and resulted in the generation of initial hypotheses. Quantitative analysis began with the coding of the data into numeric values corresponding to the study variables. Then various statistical techniques were used to discover relationships between the variables. This analysis was guided initially by the hypotheses generated by the qualitative analysis. Finally, qualitative analysis was also used after the initial statistical results were generated in order to help clarify and explain them. All of these techniques are discussed briefly in the following sections and are elaborated during the discussion of results.

Data Collection

The data for this study was collected between December, 1995, and April, 1996. The data collection procedures included gathering official documents, participant observation [14], and structured interviews [10].

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 [10]. Some of the documents which provided data for this study were:

- organizational charts
- process descriptions
- inspection data collection forms
- online newsgroup

Much of the data for this study was collected during participant observation of 23 inspection meetings. During the observations, the observer collected data on the lengths and topics of discussions, but did not play a direct role in the inspection process.

The other important data source was a set of interviews conducted with inspection participants. These interviews were semi-structured; each interview started with a specific set of questions, the answers to which were the objective of the interview. However, many of these questions were open-ended and were intended for (and successful in) soliciting other information not foreseen by the interviewer.

Data Analysis

Initial qualitative analysis on the data began about halfway through data collection. The first analysis was similar to the "constant comparison method" described by Glaser and Strauss [7] and the comparison method suggested by Eisenhardt in [4]. The method consisted of a case-by-case (meeting-by-meeting) comparison in order to reveal patterns among the characteristics of inspection meetings. The goal of this initial analysis was to suggest possible relationships between variables. These suggested relationships would then be further explored quantitatively where appropriate.

The quantitative variables chosen for analysis fall into three categories. First are the dependent variables, all of which have to do with the time or effort spent in the inspection process. Secondly, there is a set of independent variables which represent the issues of interest for this study, i.e., organizational issues. Finally, there are two intervening variables, size and complexity of the inspected material. These variables must be taken into account so that the relationships between independent and dependent variables will not be masked by them.

The quantitative analysis used in this study was fairly simple and straightforward. We began by looking at descriptive statistics (mean, minimum, maximum, median, standard deviation) for each of the variables. This helped to form an overall picture of the scope and shape of the data. Then we calculated Spearman correlation coefficients to determine which variables were statistically related (especially which organizational characteristics were related to measures of communication effort).

Qualitative data and findings were also used to help illuminate and explain the statistical findings. This was done in a more ad hoc way, by simply searching the field notes for anecdotes or quotes which shed some light on a particular finding. As well, after the initial quantitative results were generated, they were presented to several key developers on the project. This technique is called *member checking* [10]. The developers' responses to and explanations for those results were recorded qualitatively and also helped illuminate the statistical results. There are several examples of this in the next section.

RESULTS

Figure 2 depicts the network of relationships between factors that affect meeting length, according to the findings of this study. Each box represents a study variable or some other factor that became relevant during the course of the analysis. Each arrow represents some sort of relationship between two variables (e.g. a correlation) that was found in the data. We will discuss these variables and relationships in detail in the following subsections.

Components of Meeting Length

Although this study used dependent variables reflective of all parts of the inspection process, this paper presents results relating only to the inspection meeting itself, in particular the length of the meeting. Besides the actual meeting length, other relevant variables break the meeting length down into the time spent in various types of discussion. All of these variables are measures of communication effort because they describe the effort expended during the meeting, which was entirely a communication activity.

The *defect* discussion time associated with an inspection consists of time taken by raising, recording, and discussing actual defects. *Global* discussion time includes discussion of issues that are applicable to other parts of the system as well as the code being inspected. Since this includes the raising and discussing of "global" defects, this category overlaps with the defect discussion category. *Unresolved* discussion time refers to discussion of issues which could not be resolved during the meeting. *Administrative* time includes time spent in administrative activities as well as the discussion of administrative or process issues. *Miscellaneous* discussion time includes miscellaneous discussions of a technical nature, including raising and discussing questions about the code being inspected which are not determined to be defects. Aside from the overlap between "defect" and "global" discussion time, the categories are mutually exclusive.

The meeting time devoted to each discussion type is strongly correlated with meeting length, but defect discussion time is the most strongly correlated. Not only does the amount of time spent discussing defects increase for longer meetings, but so does the percentage of time spent in discussion of defects. In other words, much of what makes a long meeting long is due to extra time spent discussing defects.

However, other discussion types also play a role in determining the length of an inspection meeting. The amount of time spent discussing unresolved and global issues increases for longer meetings, as does the percentage of meeting time devoted to such discussions. Miscellaneous discussion time also seems to account for a considerable amount of the extra time spent in longer meetings.

Relatively speaking, the time spent in administrative tasks in an inspection meeting stays fairly constant and is nearly independent of the meeting length.

Organizational and Other Factors

To determine which organizational characteristics are relevant with respect to the amount of time spent in different types of discussion (and thus to the overall length of the meeting), we examined relationships between variables statistically using Spearman correlation coefficients. This test was chosen because it is non-parametric and does not require that the underlying distributions of the variables be normal. To examine the effect of the intervening variables, we also conducted the same tests between variables after subsetting the data by size and complexity. This was done to see whether or not certain relationships existed, not for all the inspections, but only for inspections of material of a certain size or complexity.

The objective of this study was to generate theory, not test it. The presentation of results below is summarized periodically with the hypotheses generated by the study findings.

Defect Discussion Time

The amount of time spent discussing defects during an inspection meeting is usefully broken down into two components. First of all, as might be expected, the defect discussion time is closely tied to the number of defects reported (Spearman coefficient 0.93, $p<.001$). However, there is some variation in the "defect discussion duration", which is the average amount of time spent discussing each defect raised in a meeting. It is useful to separate these two factors because the data shows that they are affected by different variables.

Data on the number of defects reported came from copies of the *Inspection Data Collection Form* for each inspection observed. These forms included a lot of other information about the inspection, most of which had already been collected during observations, so the forms served as a validation instrument.

It is somewhat surprising that the number of defects reported in a meeting is statistically unrelated to the size of the material being reviewed. *Size*, one of the two intervening variables used in this study, was coded into a three-level ordinal variable for analysis purposes. Size information was also provided on the Inspection Data Collection Forms.

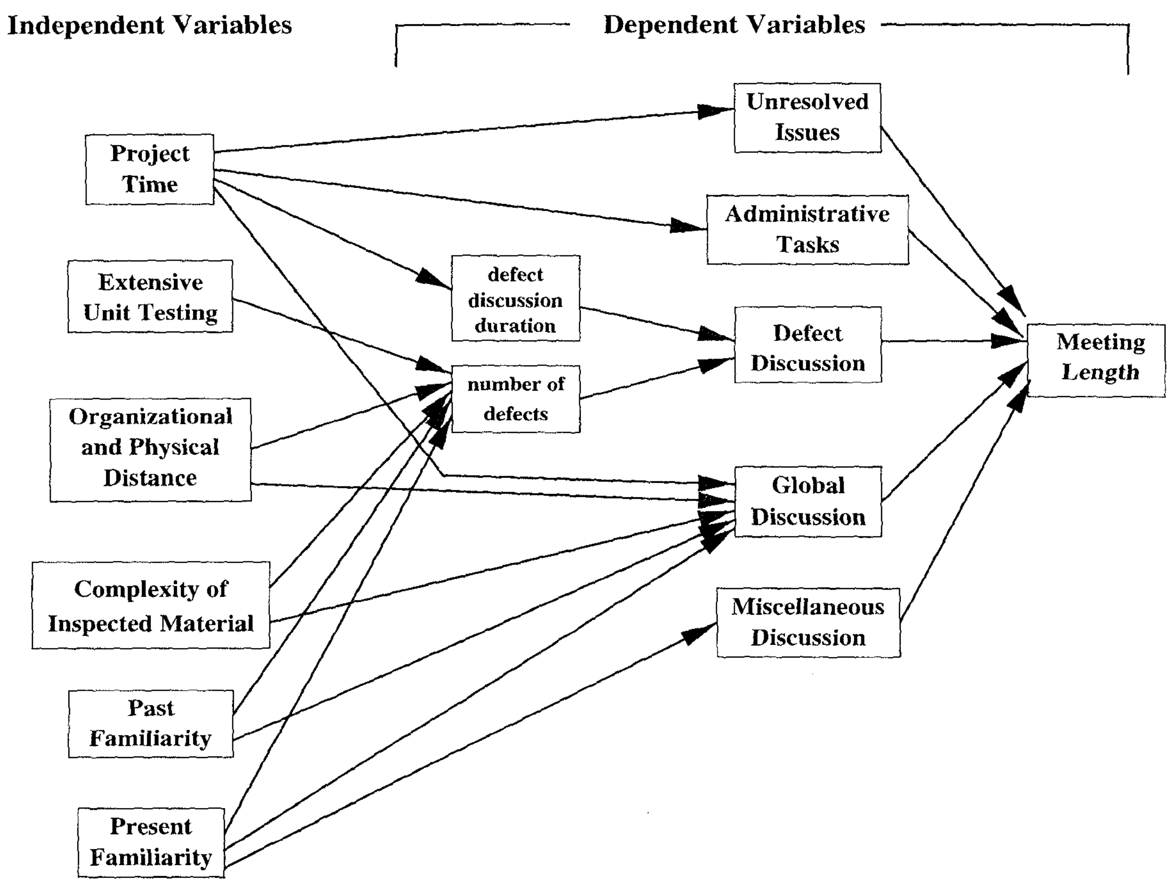

Figure 2: The network of relationships between variables.

The other intervening variable in this study, *complexity* of the inspected material, did seem to have a moderate effect on the number of defects reported. Fewer defects were reported when complexity was high (Spearman coefficient -0.5, p<.05). It might be reasonable to assume that material of high complexity actually contained fewer defects (maybe because it was assigned to more skilled developers). However, another explanation is that complex code was not inspected as carefully as less complex code. Fewer defects may have been reported because inspectors had to spend more time understanding the code and thus did not have adequate time to search for defects. Complexity was originally coded on a five-point subjective scale (based on interview data, described below), but was collapsed down to three levels for analysis.

> **Hypothesis:** The more complex the material is, the fewer defects will be reported.

Under certain conditions, fewer defects tended to be reported when the inspection participants were more familiar with each other. Two measures of familiarity were used in this study, both based on pairs of inspection participants, and both ratio-valued. *Present familiarity* reflects the degree to which the participants in an inspection interact with each other on a regular basis, and thus share common internal representations of the work being done. The value of this variable is the proportion of *pairs* of participants in the set of inspection participants who interact with each other on a regular basis. *Past familiarity* reflects the degree to which a set of inspection participants have worked together on past projects. This is assumed to contribute to a shared internal representation, not of the current work, but of the application domain in general, and a shared vocabulary. Past familiarity represents the percentage of pairs of participants who have worked together on past projects. Both types of familiarity measures are based on information gathered during interviews with each project member.

When the material being inspected was of low complexity, fewer defects were reported when the inspection participants were very familiar with each other, based on either past or present working relationships (Spearman coefficients -0.95 and -1, respectively, p<.1). Also, no matter what the complexity, fewer defects were reported when the inspection participants were familiar based on past working relationships and the material inspected was small in size (Spearman coefficient -0.87, p<.1). So, for some types of material, closer past or present working relationships between the inspectors results in fewer defects reported. This may indicate that developers are reluctant to report all the defects they find in material authored by close colleagues, or that they tend not to inspect such material as carefully as that authored by developers who are less familiar to them. Yet another possible explanation is that the familiarity measures also reflect the average experience of the inspection participants. That is, people who have been working (in the company or on the project) longer will be more familiar with more people. Thus the fewer number of defects is actually a result of experience, not familiarity. However, no significant correlations were found between the familiarity measures and a rough measure of experience which was formulated for this purpose.

> **Hypothesis:** The more familiar the inspection participants are with each other, the fewer defects will be reported.

Familiarity information was collected through interviews. Each interview used a tailored interview form, or guide [14], which included the questions and topics to be covered. These included information missing from the data form for a particular inspection, questions about organizational relationships, data on inspection activities other than the meeting, and information on the code inspected. These forms were not shown to the interviewee, but were used as a checklist and for recording answers and comments. In some cases, the more straightforward questions were asked via email. This was requested by the project management to reduce the amount of time the project personnel had to spend in interviews. Most interviews were audiotaped in their entirety. Extensive field notes were written immediately after each interview. The tapes were used during the writing of field notes, but they were not transcribed verbatim.

Another indication of familiarity is whether or not the author was in the "core group" which, for the purposes of this analysis, is defined as the eight developers who interact with other developers the most. This group consisted entirely of CSC developers, including the two CSC technical leads. All of the inspections included participants in the core group, but very few inspections involved exclusively core group members.

Inspections with a core-group author had less than half the number of reported defects than those with non-core group authors. And when the author was one of the technical leads, the number of reported defects was less than a third than in other inspections. One developer explained this latter result by explaining that one of the technical leads is very experienced, and the other, although not very experienced, is "a whiz". The classes assigned to the technical leads also tended to be lower in complexity than other classes as well.

There is also some evidence that extensive unit testing prior to the inspection reduces the number of defects reported in the meeting, which is intuitively logical. In two inspections, such extensive testing took place. In one of the inspections, one defect was reported, and two were reported in the other (much lower than the average of about 9). The low defect level cannot be explained by size or complexity. Because there were so few defects, there was very little defect discussion time, and the meetings themselves were correspondingly short. This result was especially satisfying to the developers to whom it was presented. Two developers, who both had leadership roles on the project, expressed the opinion that unit testing was a vital part of the development process, and this result was an indication that it was effective. However, since we do not know the actual defect densities of these classes, we might conclude that the inspectors may not have inspected as carefully because they knew that the classes had had extensive unit testing.

> **Hypothesis:** When more unit testing is performed prior to the inspection, fewer defects are reported.

There is also evidence that organizational and physical distance have an effect on the number of defects reported. *Organizational distance* refers to the degree of management hierarchy between members of the organization. In this study, each inspection was either organizationally "close" (all the participants reported to the same CSC manager) or organizationally "distant" (at least one participant from FDD was present). *Physical distance* reflects the number of physical boundaries between the inspection participants. In this study, physical distance takes on three values, corresponding to a set of inspection participants with offices on the same corridor, in the same building, or in separate buildings. The data used to evaluate the distance measures was collected during the prior ethnography phase, and was stored in the AD model built during that phase.

Both organizational and physical distance played a role in one particular inspection with respect to the number of defects reported. This inspection meeting was an outlier, the longest in the data set, at 100 minutes. The author was an FDD developer, while all the inspectors were from CSC (this was an unusual situation in the data set). Consequently, the inspectors were not very familiar with the class before they had inspected it in preparation for the meeting. This meeting also had the highest number of defects reported in the data set, 42. This may have been partly a direct result of the high organizational and physical distance between the participants, particularly the author. Fourteen (compared to an average of 2) of the defects were global in nature, meaning that they were defects that had been raised in previous inspections. However, the author of the outlier inspection was physically and organizationally removed from the participants in those previous inspections. This may have contributed to a lack of communication about the global defects. This is consistent with remarks from developers, who described developers in other parts of the organization as "isolated".

> **Hypothesis:** The closer the inspection participants are, either physically or in the reporting structure, the fewer defects will be reported.

The other factor contributing to the amount of time spent discussing defects, besides the number of defects reported, is the defect discussion duration, or the average amount of time spent discussing each defect. The defect discussion duration, surprisingly, is unrelated to either size or complexity of the inspection material. In fact, it is not correlated, in general, with any of the study variables. Significant correlations were found only under certain conditions. For example, for material of medium complexity, the duration of *global* defect discussions decreased over time. That is, the later in the project that the inspection occurred, the less time was spent discussing each global defect (Spearman coefficient -0.81, $p<.05$). As discussed in the next section, this most likely has more to do with the global nature of those defects than any property of defects in general.

In summary, a large part of the variation in meeting length is accounted for by the amount of time spent discussing defects, which in turn is largely dependent on the number of defects reported. This finding is somewhat comforting because the main purpose of an inspection meeting is, usually, to discuss defects. The number of defects is related to nearly all of the study variables, under different circumstances. However, the defect discussion duration also plays an important part in the amount of meeting time spent discussing defects. Unlike the number of reported defects, the defect discussion duration does not seem to be affected in general by any of the organizational variables, but under certain conditions it seemed to decrease over the course of the project. It should be noted that defect data was not available for 6 of the inspection meetings observed. The findings related to the number of defects or the defect discussion duration are based on only 17 inspection meetings, instead of the 23 that comprise the whole dataset.

Global Discussion Time

The time spent discussing global issues (including global defects) in an inspection meeting was strongly affected by a number of factors, as can be seen from the proliferation of arrows pointing to it in Figure 2.

First of all, global discussion time tended to be lower when the inspection participants were very familiar with each other, based on past working relationships. This correlation was not particularly strong in general (Spearman coefficient -0.38, $p<.1$), but was stronger for inspections of small amounts of material or material of low complexity. Also for material of low complexity, there was a strong tendency for global discussion time to be low when the inspection participants currently worked together a great deal (i.e. when present familiarity was high, Spearman coefficient -0.9, $p<.05$). In other words, people who interact on a regular basis spend less time discussing global issues only when the material being inspected is not very complex, but past working relationships have a more general effect. One developer addressed the latter result by observing that coding standards (which were the subject of many of the global discussions) are similar on all projects at CSC. So people who have worked together on past projects have most likely worked through some of these global issues together before, and thus it takes them less time to discuss them in the present. Also, it may be that developers are likely to discuss such issues outside the meeting with inspectors with whom they have worked before, thus reducing the need to discuss them during the meeting.

> **Hypothesis:** The more familiar the inspection participants are with each other, the less time they will spend discussing global issues.

There were some very specialized relationships between global discussion time and organizational and physical distance in some parts of the data.

For material of low complexity, there was strong tendency for more time to be spent discussing global issues when the inspection participants were organizationally or physically distant (Spearman coefficient 0.87, $p<.1$). However, the effect of organizational distance on global discussion time is very different when we restrict the data to inspections of large amounts of material. For such inspections, there was *less* global discussion time when the participants were organizationally distant (Spearman coefficient -0.64, $p<.1$). That is, more organizationally distant inspection participants spent less time on global issues when inspecting large amounts of material. These results are contradictory, and they imply that any effect that organizational distance has on the amount of global discussion time is overshadowed by the size and complexity of the material to inspect. It may be that large size, at least in some cases, leaves little time for inspectors to spend on "cosmetic" defects, which are often global. On the other hand, low complexity may allow inspectors to spend more time on such defects.

> **Hypothesis:** The closer the workspaces of the inspection participants, physically, the less time they will spend discussing global issues.

> **Hypothesis:** The distance between inspection participants in the reporting structure has an effect on the time they will spend discussing global issues, but depends on the size and complexity of the material being inspected.

This low complexity case is illustrated with the outlier meeting mentioned earlier (the longest meeting, at 100 minutes). The distance measures for this meeting were high, and it also included a large amount of global discussion. Global discussion constituted 18 minutes of the inspection meeting, which was much higher than the average of about 4 minutes. The complexity of the material was low, and it was small in size. The major factor seemed to be the organizational and physical distance of the author. Below is an excerpt from the field notes:

> One of the reasons this inspection was so long was that every "global" issue that had been hashed over in previous inspections was hashed out here as well, even a lot of things that had already been taken care of in [the code generator]. However, they all seemed to be a surprise to [the author], who hadn't gotten any of this presumably because he's at [FDD].

In some of the results above, the complexity of the material being inspected played a role by determining the conditions under which some results held. But complexity also had a direct relationship with global discussion time in the dataset as a whole. In general, the more complex the material, the less time was spent discussing global issues (Spearman coefficient -0.58, $p<.005$). This may indicate that, with highly complex material, the available time was spent discussing weightier issues than global defects, which are often "cosmetic".

> **Hypothesis:** The more complex the material being inspected, the less time will be spent discussing global issues.

Global discussion time also decreased over time to some extent, especially for material which was large or highly complex (Spearman coefficient -0.53, $p<.001$). This was explained by one developer as largely due to the role of the code generator, which was being developed concurrently. Many of the defects which were raised repeatedly in different inspections (i.e. global defects) were eventually remedied by implementing the fixes into the code generator. So, early in the project, a lot of effort was made to specify these problems and solutions carefully for the developers of the code generator, so that they would be implemented correctly.

> **Hypothesis:** The later in the project the inspection occurs, the less time will be spent discussing global issues.

Other Discussion Types
Miscellaneous discussion time does not decrease significantly over time, nor is it significantly related to size or complexity. However, one component of miscellaneous discussion time (the amount of time spent asking and answering questions about the code being inspected) tends to be lower when the inspection participants are familiar, based on present working relationships (Spearman coefficient -0.65, $p<.1$). As explained by one developer, people who work together a lot are simply used to communicating, so can relay ideas very quickly. They also tend to discuss many issues outside the meeting, so less time is spent on them in the meeting.

As mentioned earlier, the time spent in administrative tasks in an inspection meeting is relatively constant, regardless of the meeting length. However, time spent in administrative tasks did decrease over time (Spearman coefficient -0.52, $p<.05$), especially for inspection material of low complexity. This is largely due to the fact that much of the administrative time in early inspection meetings was spent in asking and discussing questions about the inspection process itself. Inspections were just beginning on this project, the inspection process document had just been released, and inspections were being performed differently for this project in several ways. Inspection process questions consumed up to 5 minutes of each inspection meeting of the first 10 (out of 23) inspections observed. After that, process questions did not arise, and the administrative procedures became a "habit", as one developer put it. Even with the extra time in the early inspections, however, differences in administrative time between inspection meetings does not account for very much of the variance in meeting length.

In general, more meeting time was spent on unresolved issues early in the project than later (Spearman coefficient -0.49, $p<.05$). This was because, as one developer explained, developers at first made an effort to resolve every issue during the meeting, even if they eventually found they couldn't. However, they later came to recognize more quickly which issues were best referred to someone else.

CONCLUSIONS

This paper describes an empirical study of code inspection meetings in a NASA-sponsored software development project. The relevant variables in this study were process communication effort (in particular the effort expended in inspection meetings, in general and in discussions of different types) and characteristics of the organizational structure (reporting relationships, familiarity, physical proximity). We found that several organizational characteristics have an effect on the amount of time spent in different types of discussions during inspection meetings. Below, we present our findings in the form of testable hypotheses, which are the main contribution of this work (these are also presented graphically in Figure 2).

First, we presented results that showed that two of the major factors that make longer inspection meetings longer are the time spent discussing defects and the time spent discussing global issues. Furthermore, the time spent discussing defects is mostly determined by the number of defects reported during the meeting. The following hypotheses (similar to those presented previously) represent the study findings which relate to factors affecting the number of defects reported:

- **H1** The more the inspection participants interact with each other on a regular basis, the fewer defects will be reported.
- **H2** The more the inspection participants have worked together in the past, the fewer defects will be reported.
- **H3** The more closely related the inspection participants are in the reporting structure, the fewer defects will be reported.
- **H4** The closer the workspaces of the inspection participants are, physically, the fewer defects will be reported.
- **H5** The more complex the material is, the fewer defects will be reported.
- **H6** When more unit testing is performed prior to the inspection, fewer defects are reported.

Except for the last two of the above hypotheses, all of these point to the conclusion that developers will report fewer defects in material authored or inspected by other developers with whom they are "close" (in terms of organizational distance, physical distance, or familiarity). Unfortunately, this finding cannot be fully interpreted without knowing more about the error histories of the classes inspected. That is, we cannot know whether those classes which had fewer reported defects actually had fewer defects, or whether the closeness of the inspection participants influenced the inspectors to find or report fewer defects than actually existed. A follow-up study of testing data could provide the insight necessary to address this question. It is important to look at this issue closely because the number of reported defects appears to have a very strong influence on meeting length. In fact, aside from the various types of discussion times, it is the only variable that is strongly and directly associated with meeting length. Thus it is important to know what factors affect the number of defects reported, besides the actual number of defects in the code.

For example, suppose we extrapolate the above general conclusion (close inspection participants report fewer defects) to imply that close inspection participants report a lower percentage of the defects that actually exist in the code. This is as reasonable a statement as any, as we have no reason to assume that the distribution of defects in classes inspected by a close group is any different from that of other classes. This would indicate that, while choosing a close set of inspection participants would seem to make the inspection meeting more efficient, it would seriously degrade its effectiveness.

Another factor in the length of inspection meetings is the time spent discussing global issues, or those issues that arise repeatedly and are relevant to the system as a whole, not just the code being inspected. This study indicated that the time spent discussing such issues is strongly related to the organizational relationships between inspection participants, as detailed by these hypotheses:

- **H7** The more the inspection participants have worked together in the past, the less time they will spend discussing global issues.
- **H8** The more the inspection participants interact with each other on a regular basis, the less time they will spend discussing global issues.
- **H9** The distance between inspection participants in the reporting structure has an effect on the time they will spend discussing global issues, but depends on the size and complexity of the material being inspected.
- **H10** The closer the workspaces of the inspection participants, physically, the less time they will spend discussing global issues.
- **H11** The later in the project the inspection occurs, the less time will be spent discussing global issues.
- **H12** The more complex the material being inspected, the less time will be spent discussing global issues.

In general, it can be concluded that inspection participants who are "close" spend less time discussing global issues. This is likely due to several factors, including the amount of discussion which goes on outside the inspection meeting, the shared vocabulary that arises from familiarity which facilitates communication, and a shared understanding of the actual issues that come up repeatedly. Because less time is spent in global discussion, a close group of participants also results in a shorter meeting. This says nothing, however, about the effectiveness of such a meeting.

These hypotheses could all be tested in carefully controlled experiments that are designed for that purpose. The study described here provides some evidence of their validity.

This study peels back just one layer of understanding about the role organizational structure plays in the efficiency of inspection meetings. Many other, deeper, questions remain, however. For example, what makes an inspection efficient? Is an efficient inspection meeting necessarily shorter? Does it have less discussion of some types and more of another? The answers to questions like these lie, at least in part, on the goals and objectives of inspection meetings, which vary from project to project. If they can be answered for a particular project, then studies like the one described here can provide guidance as to the organizational factors which can be manipulated to meet the project goals.

Some of the qualitative data in this study indicated the complexity of these underlying questions. In the outlier meeting mentioned earlier, for example, recall that the number of defects reported was very high and the author was organizationally and physically distant from the other participants. He had not interacted with the inspectors during implementation of that class. This suggests the following argument. Different developers may be sensitive to different types of code errors, depending on their experience. The developers with whom an author consults during development, then, will help to eliminate certain types of errors from that author's code. If those same developers are those who inspect that code, they may not find many errors because those they are most aware of have already been eliminated. But if a different set of developers inspects the class, then they may bring different sensitivities to the inspection and thus find other errors (although they may take longer to do it). This may be what happened during the long outlier inspection. One developer addressed this very issue during an interview:

> She can imagine that if the inspectors are the same people who helped craft the code, then they're not likely to find anything wrong with it. So this may be a reason to choose inspectors that are not that familiar with the code.

The above anecdote is meant simply to underscore the fact that the work described in this paper helps to enable a whole area of research. Further work in the effects of organizational structure on the productivity of development processes has potential for profoundly influencing the success of software development projects. This study not only illustrates one effective way of conducting such investigations, but also provides some hypotheses with which to begin.

ACKNOWLEDGEMENTS

The authors would like to thank the members of the development team on the project used for this study for their patient cooperation. We would also like to thank the Experimental Software Engineering Group at the University of Maryland for their input at various stages of this work, including review of drafts of this paper. Also helpful were the comments from anonymous ICSE reviewers on the original submitted version of this paper.

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(11), Nov. 1988.

[2] K. M. Eisenhardt. "Building Theories from Case Study Research". *Academy of Management Review*, 14(4):532–550, 1989.

[3] J. R. Galbraith. *Organization Design.* Addison-Wesley, 1977.

[4] J. F. Gilgun. "Definitions, Methodologies, and Methods in Qualitative Family Research". In *Qualitative Methods in Family Research.* Sage, 1992.

[5] B. G. Glaser and A. L. Strauss. *The Discovery of Grounded Theory: Strategies for Qualitative Research.* Aldine Publishing Company, 1967.

[6] H. Krasner, B. Curtis, and N. Iscoe. "Communication Breakdowns and Boundary Spanning Acitivities on Large Programming Projects". In G. Olsen, S. Sheppard, and E. Soloway, editors, *Empirical Studies of Programmers, second workshop*, chapter 4, pages 47–64. Ablex Publishing, New Jersey, 1987.

[7] R. E. Kraut and L. A. Streeter. "Coordination in Software Development". *Communications of the ACM*, 38(3):69–81, Mar. 1995.

[8] Y. S. Lincoln and E. G. Guba. *Naturalistic Inquiry.* Sage, 1985.

[9] J. March and H. A. Simon. *Organizations.* John Wiley, New York, 1958.

[10] D. E. Perry, N. A. Staudenmayer, and L. G. Votta. "People, Organizations, and Process Improvement". *IEEE Software*, July 1994.

[11] S. J. Taylor and R. Bogdan. *Introduction to Qualitative Research Methods.* John Wiley and Sons, 1984.

[12] E. Yu and J. Mylopoulos. "Understanding "why" in software process modeling, analysis, and design". In *Proceedings of the 16th IEEE International Conference on Software Engineering*, Sorrento, Italy, 1994.