Measurement Programs in Software Development: Determinants of Success

Anandasivam Gopal, M.S. Krishnan, Tridas Mukhopadhyay, and Dennis R. Goldenson

Abstract—Measurement programs in software organizations are an important source of control over quality and cost in software development. The findings of this research presented here are based on an industry-wide survey conducted to examine the factors that influence success in software metrics programs. Our approach is to go beyond the anecdotal information on metrics programs that exists in the literature and use the industry-wide survey data to rigorously test for the effects of various factors that affect metrics programs success. We measure success in metrics programs using two variables—use of metrics information in decision-making and improved organizational performance. The various determinants of metrics program success are divided into two sets—organizational variables and technical variables. The influence of these variables on metrics programs success is tested using regression analysis. Our results support some of the factors discussed in the anecdotal literature such as management support, goal alignment, and communication and feedback. Certain other factors such as metrics quality and the ease of data collection are not as strongly influential on success. We conclude the paper with a detailed discussion of our results and suggestions for future work.

Index Terms—Software metrics programs, empirical methods, survey, regression analysis, software development, measurement programs, determinants of success, software engineering.

1 Introduction

Software metrics are increasingly being considered a vital part of software engineering as the software industry grows. There is a great deal of literature in this area that discusses the relevance and importance of measurement and its contribution to improved quality and productivity. Software metrics are initiated with the belief that they will improve software engineering and management practices. The rationale arises from the notion that you cannot improve something without first measuring it. Software metrics and measurement activities are, thus, organization-wide attempts to institutionalize the concept of measurement and feedback. Measurement and metrics teams are therefore the first link in an overall process-based improvement effort in a software development environment. The basic rationale behind investments in metrics is that metrics programs, once implemented and utilized, should lead the software organization toward more disciplined processes through an efficient use of the feedback mechanism.

A software metric is defined as a method of quantitatively determining the extent to which software process, product, or project possesses a certain attribute [7]. Since the software industry is plagued with productivity and quality issues, software engineers and managers recognize the need to better understand and, hence, better manage the software development process and be able to make the necessary changes to improve productivity and quality. In order to do this, metrics relevant to software engineering processes and products have to be collected and analyzed. Based on the feedback of this analysis, engineers and managers can make necessary changes to either the process, product, or both. Such improvements should result in increased productivity and quality, reduced cycle time, and lesser costs in the long run. Case studies in the literature show some evidence of such improvements occurring through the deployment of successful metrics programs [7], [8], [23].

As noted above, although some evidence in support of metrics does exist, many companies find software measurement to be a complex and difficult undertaking. One study showed that fewer than 10 percent of the industry classified metrics programs as positive [7]. Another study reports that two out of three metrics initiatives did not last beyond the second year [23]. In a recent case study on metrics implementation, Iversen and Mathiassen identify that the implementation of the metrics program is long and complex and the involved actors are confronted with a number of dilemmas based on contradictory demands and value conflicts related to the metrics program [12]. Thus, it is important from a practitioners' viewpoint to examine, in more detail, the interrelationships of various factors that have been theorized to affect success in metrics programs implementation. One of our primary goals in this paper is to identify the various success factors on metrics programs success and study their direct and indirect effects on metrics programs success.

Prior literature in the area has indicated that there are several factors that influence the success of metrics programs in software development. However, the evidence presented in the literature is mostly anecdotal or based on a single case study. There has been limited rigorous empirical research in identifying which factors are more influential in

- A. Gopal is with PRTM, 1050 Winter St., Waltham, MA 02451. E-mail: agopal@prtm.com.
- M.S. Krishnan is with University of Michigan Business School, 701 Tappan St., Ann Arbor, MI 48109-1234. E-mail: mskrish@umich.edu.
- T. Mukhopadhyay is with the Graduate School of Industrial Adminstration, Carnegie Mellon University, Pittsburgh, PA 15213. E-mail: tridas@cmu.edu.
- D.R. Goldenson is with the Software Engineering Institute, Carnegie Mellon University, Pittsburgh, PA 15213. E-mail: dg@sei.cmu.edu.

Manuscript received 29 Mar. 2001; accepted 2 Aug. 2001.
Recommended for acceptance by S. Pfleeger.
For information on obtaining reprints of this article, please send e-mail to: tse@computer.org, and reference IEEECS Log Number 110225.

the success of such programs. Our work fills this gap by examining metrics programs success through the use of empirical data rather than individual case studies. Our findings are based on data collected from a survey that was conducted between December 1998 and February 1999 under the auspices of the Software Engineering Institute at Carnegie Mellon University. Our analysis corroborates some of the anecdotal evidence from the literature and also provides some new insights for understanding the relative influence of different factors on the success of a metrics program.

The contributions of this paper are multifold. First, unlike case studies on metrics programs in a single organization, we study the issue of metrics programs success and their causal factors using data collected from managers in multiple software firms. This approach enables us to generalize our conclusions better than in-depth case studies. We believe that our analysis complements the findings of case studies and the combined results will provide the practitioner with a more complete picture of metrics programs. Second, we operationalize success in metrics programs through the use of two new variables instead of more traditional measures such as longevity or budget allocated to these programs. The two variables of metrics program success used in our analysis more directly address the business value returns from a metrics program. Third, we analyze the direct and indirect effects of some of the critical factors that have been theorized to affect metrics programs on metrics programs success. We believe this two-level analysis will provide new insights to clarify some interrelationships between the critical factors and metrics programs success for researchers and practitioners alike.

The rest of the paper is organized in the following way: In the next section, we review the prior literature. In Section 3, we discuss our model and identify various factors used to analyze metrics programs. In Section 4, we describe variable definitions and our research methodology. Section 5 discusses the results from our statistical analysis and Section 6 concludes the paper with some directions for future research.

2 Prior Literature

The literature in the area of software metrics has been almost entirely based on case studies and qualitative analyses of metrics programs in some organizations. Case analyses typically include descriptions of the introduction of certain kinds of metrics in some organizations, their salient features, and the pitfalls that a metrics manager must look out for. Although some evidence exists to show that metrics programs are difficult to introduce and maintain, no rigorous empirical research has been carried out in finding out what factors are influential in determining the success of such programs. One of the reasons for the gap in this literature is the lack of data on metrics programs across organizations. Data on metrics programs are difficult to obtain and are often subjective in nature, thereby making analysis difficult. Moreover, metrics programs differ in organization and structure across different software firms. This makes it more difficult to conduct any kind of quantitative analysis on data obtained from these firms. Therefore, case studies become the source of evidence. We next review some of the prominent studies in this literature.

A good example of a case discussion of metrics program implementation can be seen in Daskalantonakis [7] who discusses the introduction and implementation of a metrics program at Motorola in some detail. He also uses his experience with the program to describe certain important factors that affect the success of the metrics programs. He focuses on the utility/usefulness of metrics chosen and the impact of stakeholder support in setting up a successful program. In a similar paper, Pfleeger [23] discusses the introduction of a metrics program implemented in Contel Corporation and the salient features of the program. The author ends the paper with a list of "lessons" learned from the introduction of the metrics program there. More recently, Iversen and Mathiassen studied the metrics implementation in the software subsidiary of a large financial institution [12]. They adopted a two-stage approach to understand the complexity in the metrics implementation process. In the first stage, they view metrics implementation as a rational, engineering process in which a metrics program is constructed and put into use. In the second stage, they treat the metrics program as an evolutionary process in which the basic values of the software organization are confronted and transformed. Related case studies can also be found on metrics programs at Eastman Kodak [27] and in the US Army [8].

Using a different approach, Hall and Fenton [10] analyze two different metrics programs in two different organizations. One of these metrics programs was successful while the other was not. The authors use a set of factors gleaned from the literature in their analysis. Their analysis shows the importance of a well-planned metrics framework, the involvement of developers and management in setting up the program, and the role of feedback and communication in ensuring success. Jeffery and Berry [13] propose a framework for evaluating the success of the metrics programs. They consider four aspects of metrics programs—Context, Inputs, Process, and Products. *Context* refers to the environment in which the metrics program is developed. *Inputs* are factors or resources that are applied to the metrics program. *Process* is the method used to develop, implement, and maintain the program, and *Products* are measures taken, reports produced, and other outputs of the program. They analyze three metrics programs using this framework. They concluded that all of the four aspects are important in setting up a metrics program, and identify the need to study the relative impacts of the four aspects in software organizations through empirical research. In a research report on a recent real-world software improvement exercise, Mayes underscores the importance of careful implementation of metrics programs and the need to build credibility around metrics for the success of software process improvement [19]. We next discuss other research streams in the literature that present theory and research methodology that may be useful in studying the complex process of software metrics implementation.

One source of theory and methodology that seems to parallel software metrics implementation and the acceptance of metrics culture in the organization is the literature on the diffusion of innovations. Innovation diffusion studies examine the rate of diffusion of new and innovative practices across and within organizations. These studies also examine factors that affect the rate of diffusion across organizations. Unlike prior research on metrics programs, there is some empirical research in this area that can be leveraged in our context. Additionally, the methods for analyzing the impacts of different factors affecting successful implementation of such innovations have been studied in this literature, and can be used in our context. Metrics programs can be compared to an administrative innovation or process innovation in that the results of such programs are uncertain in the short run, generally positive in the long run, and their adoption is never homogenous across organizations. Although the theory of innovation diffusion may not transfer directly to metrics programs, methodological insights can be gained by a short review of this literature. Innovation diffusion models have been studied in different contexts such as hospital process innovations, electronic data interchange (EDI) adoption, and deployment of digital switches in the telecommunications industry. Next, we briefly review this literature.

Ramamurthy and Premkumar [25] study the impact of various organizational and innovation factors on the level of internal and external diffusion of EDI across companies. By diffusion, they refer to the extent to which EDI is integrated into the company's work processes. The authors use a two-stage model where a set of innovation-specific and organizational factors affect the rate of diffusion in the first stage. In the second stage, the rate of diffusion affects the organizational performance. Thus, they are able to tease out the direct and indirect effects of EDI adoption in their sample. In another paper, they study the effects of innovation characteristics such as complexity and compatibility on the level of internal and external diffusion in companies implementing EDI using multivariate regression [24].

In a different context, Majumdar and Venkataraman [17] studied the adoption of digital switches in the US telecommunications industry and the factors that affected the decision to adopt digital switches. They use data from over 10 years to analyze the influence of competitive pressures in the marketplace and the slack availability in the firm on the decision to adopt. They collapse the data from 10 years into a cross-sectional sample and use multiple regression to analyze the data. Their results show that firm-level factors are significant in determining adoption. Kimberley and Evanisko [15] studied the influence of organizational and individual factors on the adoption of technological and administrative innovations in hospitals. They used behavioral survey techniques in collecting the data through a questionnaire and used regression for statistical analysis. Their results show that organizational variables explain diffusion better than individual factors in hospitals. Further research on innovation diffusion can be found in [5] and [3].

A specific example of a process innovation in software development that can be studied from a methodological viewpoint for drawing parallels to study metrics implementation is the adoption of CASE tools. CASE tool adoption is expected to boost productivity, reduce costs, and improve quality in the long run. However, the success of CASE adoption has been mixed in the industry in spite of management support and good reviews in the practitioner press. Although adoption of CASE tools in a software organization is not the same as metrics implementation, we believe that parallels can be drawn from the research methodology used in the case adoption literature. We next discuss some research on CASE adoption from a methodological viewpoint.

A study conducted in 1992 reported that one year after introduction, 70 percent of the CASE tools were never used, 25 percent were used by only one group, and 5 percent were widely used but not to capacity [14]. Similarly, Aaen et al. [1] found that less than 20 percent of the Finnish and Danish companies in his study were close to routine users of CASE, even though 24 percent had used them for more than three years. The literature shows that CASE adoption has also been slow and plagued with difficulties. This consistent lack of success and the complexity in introducing CASE tools in a software organization had drawn the attention of researchers. We next review a few studies from this literature.

In an attempt to examine the factors that influence CASE adoption, Iivari [11] used retrospective, perceptual data from practitioners in the field to analyze factors that affect CASE usage. Innovation diffusion was used to identify a set of factors that might affect CASE usage. As noted earlier, a two-stage model was used wherein the first stage assessed the effect of critical factors on CASE usage. The second stage analyzed the effects of CASE usage on perceived improvements in productivity. Another study on CASE adoption used data from a questionnaire administered to Danish and Finnish organizations using CASE tools to examine the effect of certain factors on CASE introduction [2]. They studied organizational and tool-related factors that could have an influence on CASE introduction. Further research on CASE adoption can be found in [9].

As noted earlier, we believe that we can draw on the methodology used in this stream of research to study metrics implementation success. In both the innovation diffusion and CASE adoption literature, there are certain underlying structures to the factors that affect successful adoption and use that can be transposed to the metrics program context. There are broadly two aspects to these diffusion models. First, the organization has a significant influence on increasing adoption by instituting processes that are conducive to the innovation. These factors can be called organizational or environmental variables and are organization-dependent. They are important in setting the tone and culture in the organization and depend heavily on the interest and support of managers in the organization. The second aspect is that of technical quality of the innovation or "tool-friendliness" in the context of CASE tools. The innovation should be easy to use and should not cause a significant overhead to the organization's functioning. This is at the level of the tool/program and has a more immediate impact on the users of the tool/program. Incorporating the new practices into the organization should be done seamlessly and with involvement from different parts of the organization. We use these two central factors in formulating our model of metrics program success in the next section.

3 Proposed Model

We draw from the methodology and theory used in the innovation diffusion and related stream of literature to study the success of software metrics implementation. Our theoretical model is depicted in Fig. 1. As shown in the figure, there are primarily two types of factors that affect metrics programs in our model. The first type refers to the *organizational* context, i.e., the level of commitment that the

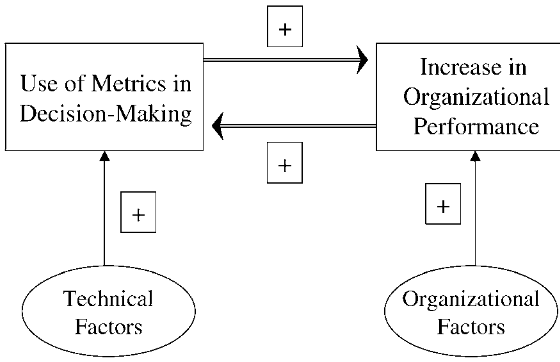


Fig. 1. Theoretical Model.

organization shows toward the metrics program. This type of factors will have a strong positive influence on the metrics programs' success and typically depend on senior management. The second category of factors is the *technical* environment, i.e., how "user-friendly" the metrics program is. The immediate environment around the metrics program should be nonintrusive and nonobstructive. The metrics program and the metrics used should be simple and straightforward and easy to use. As noted earlier, prior literature has addressed this issue at length as being one of the reasons for poor performance and termination of the program [23], [10]. Moreover, the effort required for collecting the information required for a metrics program should not add significantly to the organization's workload. These factors are referred to as technical factors in our model. We will discuss these factors in more detail below.

We hypothesize that the two types of factors described above—the organizational and technical factors, contribute to the success of a metrics program. The technical factors play a direct role in making the metrics program's information more accurate, timely, and easy to use. Moreover, the technical factors also influence the level of effort and time it takes in the organization to collect and analyze the metrics data. At the level of the smallest organizational unit that produces and utilizes metrics data, typically the project team, the technical factors contribute to easing the resistance to metrics programs and increasing the confidence and involvement of people in the organization. The organizational factors enable success by obtaining buy-in from all the parties concerned and by setting the organizational culture that recognizes and accepts metrics programs. These factors also enable metrics programs by providing valuable management support and resources for the proper functioning of the metrics program. Thus, we hypothesize that these two types of factors are important in determining the success of a metrics program.

We measure the success of a metrics program in two ways. Past studies on software metrics have quantified success of a metrics program by using either the longevity of the program or the budget allocated to it. However, we believe that these measures of success capture a limited dimension of success and, hence, may be misleading. The budget allocated to a metrics program or the number of years it has been active do not necessarily denote success if the metrics information is never used. Moreover, discussions with some project managers at the SEI have indicated that continuation of an unsuccessful metrics program is often a political decision within the organization; this is practically true in some organizations. Therefore, in our model, we consider behavioral aspects of a metrics program's output—*use in decision making* and *impact on organizational performance.* We believe that these measures are also closer to addressing business value returns from metrics programs. The first construct addresses whether the metrics program is used in the organization in making decisions. If the output of the metrics program in the organization is never used in making decisions, the subsequent benefits of a metrics program will be negligible. The second construct, increase in organizational performance, is a second-order effect of use in decision making. Increased use of metrics information in decision-making should improve organizational feedback, thereby increasing transparency into the software development process. In the greater organizational context, this should result in increases in organizational performance [6]. We feel that these two constructs capture the success of a metrics program better than just longevity or the cost of a metrics program. In subsequent discussions, we use these success factors in formulating our model.

We characterize the success of a metrics program using these two success variables through a two-stage approach. This two-stage approach characterizes the manner in which metrics programs benefits spread through the organization and is consistent with prior research methodological approach used in innovation diffusion and CASE tools [11], [24]. In the first stage, the organization should be able to use the output of the metrics program in its decision making. In the second stage, the consistent use of metrics information in decision making should lead to enhanced organizational performance. As mentioned before, one of the aims of this paper is to tease out the direct and indirect effects of certain critical factors on success. As a first step toward that aim, we hypothesize that the technical factors discussed before will have a direct influence on the use of metrics programs in decision-making. The technical factors make the metrics collection, analysis, and dissemination efficient and painless, thereby providing project personnel with incentives to use them in decision-making. Similarly, we hypothesize that the organizational factors will have a direct impact on the improvements in organizational performance. These organizational factors facilitate organizational improvements when decision-making is done on the basis of metrics information. The organizational factors also ensure that the organizational culture accepts and is built around the relevance of metrics information. Therefore, these factors in conjunction with the use of metrics information in decision-making should lead to organizational performance improvements.

We also hypothesize that the relationship between use in decision-making and increase in organizational performance is positive and is cyclical.[1] The use of metrics in decision-making will lead to increases in organizational performance, all else being equal. As mentioned above, the use of metrics information in decision-making in a systematic manner will provide project managers more

1. In statistical terms, the relationship that we postulate between the two dependent measures can be tested rigorously through time series and panel data analysis. However, since the data we have is from a survey, we do not have time-dependent data. The analysis here acts as a first step in the direction of more research in this area.

transparency into the development process. It will also provide the manager with feedback on resource allocation and utilization levels across projects, thereby enabling optimal use of resources. Although mere implementation of the metrics by itself can lead to performance improvements since it makes software developers think about their activities, we believe that without the use of metrics information in decision-making, these benefits may be minimal.

We also hypothesize that improvements in organizational performance will be associated with increased use of metrics programs in decision-making, all else being equal. This is due to the presence of a lagged effect of increases in performance due to use in decision-making. A perceived increase in performance in one time-period can be associated with increased effectiveness in managing development projects, which is associated with an increased or consistent use of metrics information in decision-making. Moreover, from a resource-sufficiency viewpoint, increases in performance could increase the resources available to all organizational units, thereby resulting in increases in resources allocated to the metrics program. However, for more resources to be allocated to facilitate use of metrics, the metrics program need to win credibility within the organization. The primary argument for improved organization performance inducing the use of metrics is the following: Once the metrics are in place and are successfully used within the organization and it leads to improved organizational performance, if the senior management believe that the metric program contributed to organizational success, then steps may be taken to facilitate further use of the metrics. This further use of metrics will in turn lead to better organizational performance and, hence, the cyclical relationship in our model.

On the basis of the above reasoning, we propose the theoretical model shown in Fig. 1. We hypothesize that the technical factors discussed above will positively influence use in decision-making and the organizational factors have a direct effect on increases in organizational performance. In addition, use in decision-making and organizational performance positively influence each other. The technical and organizational factors that we study in this paper are described below. In the interest of parsimony, we selected a subset of all the critical factors discussed in the metrics programs anecdotal literature. Given that our research design called for data from multiple organizations and was based on a questionnaire, we selected the technical and organizational factors that are the most common among software organizations and have been discussed in past research. We have also made sure that the factors chosen are conducive to survey measurement.

3.1 Technical Factors

1. *Metrics Collected*: Anecdotal case studies have emphasized the importance of the right kind of metrics that should be collected as one of the factors influencing the use in decision-making [7], [23]. An important consideration in any metrics program is the decision regarding the metrics that need to be collected. There are several kinds of metrics that can be used in decision-making. These include simple cost and schedule metrics and more sophisticated data on customer satisfaction and quality assurance. A systematic and process-oriented system of collecting metrics information will contribute to the use of metrics information in decision-making by increasing the project managers' perceptions of metrics information reliability, thereby increasing their use in decision-making. We hypothesize that the frequency with which metrics are collected will have a strong positive impact on their use in decision-making, everything else being equal.
2. *Data Collection Procedures*: Prior research has discussed the importance of nonintrusive and streamlined data collection procedures. Pfleeger discusses in detail the importance of smooth and nonobstructive data collection procedures in determining the success of a metrics program [23]. Daskalantonakis [7] and Offen and Jeffery [21] also make similar points. Optimized data collection procedures affect metrics programs in two ways. First, they do not reduce programmer productivity significantly, which typically creates resistance and reluctance to use metrics information in decision-making. Second, they increase the accuracy and reliability of the data, thereby increasing the confidence of managers who use this information in decision-making.
3. *Quality of Metrics Collected*: Most practitioners agree that the metrics collected should be accurate, relevant, and informative [10], [7]. They should convey information, and changes in the metric should have a relevant meaning in the appropriate context. The quality of data collected directly influences the level of the stakeholder's confidence in these metrics. It also increases the range of analyses that can be performed on this data, whether quantitative or qualitative. Therefore, the quality of data collected should have a positive impact on their use in decision-making.
4. *Metrics Analyses*: The rigor and usefulness of the analysis procedures carried out on the metrics data is a key factor in how the results are used. Analyses carried out could range from just maintaining archival data to making comparisons across time to carrying more in-depth analysis, like statistical analyses, experiments, and pilot studies. The more sophisticated analyses are carried out, typically the more the results of such analyses can be used to gain insights into the development process. The sophistication of the analyses should, thus, lead to increased use in decision-making, all else being equal.
5. *Training*: The availability of training on metrics and measurement is an important factor. Personnel who are trained or exposed to metrics and the potential benefits from software metrics will tend to use it more than people who have not been informed about metrics. In a recent metrics implementation study, the need to share metrics data and publicize metrics related findings and potential benefits within the organization was identified as one of the critical success factors [12].

6. *Automated Tools*: Prior literature in the area has proposed that the presence of automated tools to gather, organize, and analyze metrics should have a positive influence on metrics programs success [10], [21]. The use of automated tools reduces the overhead of data collection and also makes analysis and dissemination of feedback easier and more accurate. Additionally, the reduction of human error in data collection increases the perceptions of managers of the validity of metrics information. Therefore, we expect this variable to positively impact the use of metrics information in decision-making.
7. *Communication and Feedback*: The appropriate and timely communication of metrics information to the decision-makers is critical in this context. Without suitable feedback, several of the above-mentioned factors may not be useful. For example, the metrics team could be very well trained, have excellent automated tools for data collection, and could perform in-depth analysis on this data. However, without adequate feedback and established communication channels, managers would never use this information. The presence of a feedback loop is central to any control system and this factor measures the quality of the feedback channel for the organization. Therefore, we expect communication and feedback to enhance the use of metrics information in decision-making.

3.2 Organizational Factors

1. *Alignment of Personnel in Setting Metrics Goals*: The Goal Question Metric (GQM) paradigm in metrics postulates that there should be alignment between the various stakeholders in formulating the goals of the metrics program [4]. In other words, for a successful program, it is important that various project managers, upper-level management, and measurement specialists work together in deciding the scope and range of the program [23], [21]. Alignment leads to three subgoals that are each important to the metrics program. First, it leads to prioritization of goals within the organization and helps to scope these goals in detail. Second, it helps define the levels of resource allocation in the organization to achieve each goal. Third, it helps achieve optimal policies for the metrics programs in the organization so as to achieve these goals. It is possible that two software organizations have similar products but have vastly differing organizational goals. The presence of alignment between the various stakeholders will ensure that the organizational goals with the highest priority get achieved. Hence, the metrics program will be tailored to achieve those goals, thereby improving organizational performance. Thus, we hypothesize that alignment of different people in setting goals will be positively associated with organizational performance.
2. *Resource Sufficiency*: A metrics program needs several resources to survive. For example, adequate funding and access to literature and trained people are important resources. The lack of adequate resources can severely hamper the benefits of a metrics program. Prior research in innovation diffusion and metrics programs has discussed the importance of resources [3], [7] and we hypothesize that resource sufficiency in the organization for the metrics program will lead to improvements in performance.
3. *Management Support*: Prior literature in both the innovation diffusion and metrics program areas has indicated that this is an important variable. The presence of management support to the metrics program is a form of institutional ratification and is necessary for the survival and performance of the program. Pfleeger [23] and Offen and Jeffery [21] describe the ways in which the contribution of management support is essential for a metrics programs' success. Management support is also important in that it helps establish the culture in the organization whereby decisions made through a rational analysis of metrics data is considered appropriate. Therefore, when managers make decisions on the basis of analysis of metrics information, these decisions receive the support of upper management, thereby leading to improved organizational performance. Thus, we hypothesize that management support will lead to better organizational performance.
4. *Maturity Level*: Theoretically, the maturity level of an organization indicates the level to which there are disciplined processes for software development in the organization [22]. Mature organizations generally possess metrics programs and the benefits from these programs should be higher for them than similar programs in less mature organizations, all else being equal. Therefore, we hypothesize that mature organizations should experience greater increases in organizational performance than less mature organizations.
5. *Institutional Beliefs*: This variable is based on the prevalent set of beliefs that exist in the organization. Institutionalization is a strong force in organizations and plays a strong role in the effects of work practices [20]. Institutional theory argues that organizations function in certain ways not due to efficiency reasons but because they are influenced by accepted norms of behavior. These norms are set either by industry leaders or by organizational bodies (such as a professional association). We hypothesize that the prevalent set of institutional beliefs in the organization should be in favor of the value of metrics information. Therefore, the presence of such a set of institutional beliefs will lead to improvements in organizational performance, all else being equal.

4 Research Methodology and Data

The methodology used to collect data for this paper was a broad-based survey administered though the World Wide Web. Lists of potential respondents were obtained from the Software Engineering Institute and from two other sources. The first of these sources was a private organization that ran tutorials and courses for metrics managers, published a journal relating to metrics information, and organized

TABLE 1
Summary Statistics $N = 214$

Variable	Mean	Std Dev	Minimum	Maximum
Use in Decisionmaking	3.165	0.985	1.00	5.00
Org. Performance	3.376	0.977	1.00	5.00
Metrics_Basic	2.044	1.014	1.00	4.75
Metrics_Advanced	3.052	1.181	1.00	5.00
Data_Collection	2.792	1.180	1.00	5.00
Quality_Metrics	2.466	0.999	1.00	5.00
Analyses	3.487	1.030	1.00	5.00
Training	3.580	1.217	1.00	5.00
Communication	2.281	1.011	1.00	5.00
Feedback	2.697	1.111	1.00	5.00
Automated Tools	3.289	0.985	1.00	5.00
Align_Other	3.120	0.885	1.00	5.00
Align_QA	2.775	1.164	1.00	5.00
Resource Sufficiency	2.993	0.991	1.00	5.00
Management Support	3.011	1.302	1.00	5.00
Institutional Beliefs	3.160	0.637	1.25	4.75
CMM Level	0.442	0.497	0	1
Soft_dev	42.60	26.13	0	100

several annual conferences for metrics professionals. The second source was a division of the US Department of Defense that coordinated metrics-related activities for the Defense software divisions and software vendors. These organizations supplied us lists of practitioners who had attended one or more of their courses. Since it was important that the respondents to our survey be familiar with metrics programs, we targeted the survey at metrics professionals in the field. Our sample was selected from these lists and consists of people who were metrics practitioners, managers, or other software practitioners who have used metrics information in their organizational decision-making. The sample includes a combination of defense organizations, defense contractors, commercial software vendors, and industrial houses.

The questionnaire items for each of the higher-level constructs were created after detailed discussions with project managers and researchers at the Software Engineering Institute. The prior research in metrics programs has little empirical work and, therefore, there are no established measures that could be used for several of our constructs. The respondents were contacted by e-mail and were directed to the location of the online questionnaire. The respondents were assigned personal authentication codes to protect the integrity and security of the data. The participants were assured of data confidentiality. The survey was conducted between December 1998 and February 1999.

In our electronic survey, it is difficult to assess the accurate response rate due to the following reasons. First, some of the e-mail addresses in our sample were not current and yet others had been changed with no forwarding address. Moreover, we excluded respondents that did not work for software development firms or were employed as consultants. The approximate response rate was 55 percent, i.e., 228 completed responses. A few cases were deleted due to missing data thereby reducing the final sample size to 214. The 214 responses were collected from 126 organizations. Some organizations had multiple metrics programs in place in different parts of the organization, which were independently managed. These were counted separately in our sample. Of the 214 respondents, 130 respondents were defense contractors or defense departments. The other 84 respondents were from the commercial software development sector. There were 52 defense organizations in the sample, while the number of commercial organizations numbered 74.

Like in most survey research, our research methodology introduces certain biases that should be discussed. Our sample is not industry wide as there are several sectors of the software market that are not included. Moreover, the respondent in our sample is typically one person from the organization that has implemented a metrics program. Thus, our results could be biased by the perceptions of the person responding to the survey. However, this is a common limitation in most survey research. We reduce the impact of this by performing reliability analysis that we describe in the next section. We also randomized the order

TABLE 2a
Questionnaire Items and Factor Loadings

Use in Decision Making and Management (Use_Mgmt) How widely are software measurements actually used in making management and development decisions?	*Factor Loadings*
a) Monitoring and managing individual projects or similar work efforts	0.77
b) Use of historical data for project planning and estimation	0.77
c) Rolled up for larger organization and enterprise wide Purposes	0.71
d) For use by individual engineers, programmers and other Practitioners	0.69
e) Changes are made to technologies, business or development Processes as a result of out software measurement efforts	0.69
Organizational Performance (Perf) In your judgment, how much has the use of software measurement improved your organization's performance?	*Factor Loadings*
a) More accurate budget estimates or ability to reduce costs	0.75
b) More accurate schedule estimates or ability to reduce cycle time	0.83
c) Better adherence to customer or user requirements or improved customer satisfaction	0.82
d) Fewer software defects, faults or failures	0.88
e) Better functionality or user interface	0.82
f) Better over-all quality of products and services	0.88
g) Improved staff productivity or reduced rework	0.84
h) More informed judgments about the adoption or improvement of work processes and technologies	0.80
i) Better work processes	0.80
j) Better strategic decision-making	0.79

TABLE 2b
Questionnaire Items and Factor Loadings (cont.)

Metrics Collected To the best of your knowledge, how often are the following kinds of measurements done in your organization?	*Factor Loadings*
Metrics-Low	
a) Product Size	0.61
b) Effort	0.85
c) Cost / Budget	0.84
d) Schedule	0.81
Metrics-Hi	
a) Results of inspections and reviews	0.76
b) Other quality measures (maintainability, portability, Complexity etc)	0.76
c) Quality assurance audit results	0.79
d) Requirements stability	0.86
e) Process Stability	0.78
Data Collection Procedures (Data_Collection) How well do the following questions describe your organization?	
a) Responsibilities and procedures for recording measurement data are clearly stated and well understood	0.91
b) Our procedures for data collection and submittal are well integrated with other responsibilities and work processes	0.89
c) People consistently provide information as planned and when requested to do so	0.75
Quality of Metrics Collected (Quality_Metrics) How well do the following questions describe your organization?	
a) The data we collect are often inaccurate or incomplete (reverse scored)	0.87
b) Our measures are well-defined to accurately capture what they are meant to portray	0.88
Kinds of Analyses Done (Analyses)	
a) The measurement data that we collect are analyzed on a regular basis	0.76
b) Comparisons are regularly made between current project performance and previously established performance baselines and goals	0.74
c) Sophisticated methods of analyses are used on a regular basis	0.86
d) Statistical analyses are done to understand the reasons for variations in performance	0.84

of questions that pertain to one factor in the questionnaire, thereby reducing respondent bias further. As mentioned above, it was not possible to track nonrespondents to our survey and so at this point, we have not been able to compare our sample with the nonrespondents to check for systematic bias. However, with a reasonably high response rate of 55 percent, we believe that we have minimized the influence of bias.

The theoretical constructs discussed earlier were measured as discussed below. Each construct for the technical and organizational factors and the two dependent variables for metrics implementation success were measured using a set of two to five questions in the questionnaire. All responses to the questions were measured on a five-point Likert scale where one was Very Low and five was Very High. The individual questions that were used are shown in Table 2. In behavioral research, reliability testing is required to ensure that the theoretical constructs are measured well. In other words, the measures should have high test-retest correlation. If a construct is measured using a number of different questionnaire items, the responses to the items should show high interquestion correlation and also be correlated highly with the composite construct. We tested for the reliability of the actual measures and report Cronbach's alpha, which tests for interquestion reliability, later in this section. Generally, a Cronbach's alpha of 0.70 and higher is considered acceptable for survey research [27].

We used principal components to confirm the loadings of items on the proposed theoretical constructs. For ease of interpretation, the scores of various questions associated with a construct were then averaged to create a composite score for that construct. For example, there were 10 items in the questionnaire for measuring organizational performance. The final score for a respondent on organizational performance was an average of the 10 individual scores. The following section lists out the Cronbach's alpha for each theoretical construct and also mentions the number of questions in the questionnaire that was used in measuring that construct. The summary statistics for the composite variables are shown in Table 1 and the individual questions and factor loadings are shown in Table 2.

Success Variables (Dependent):

1. **Use in Decision making**: (*Use_Mgmt*) alpha = 0.78, five items.
2. **Increase in Organizational Performance**: (*Perf*) alpha = 0.94, 10 items.

Technical Factors:

1. Metrics Collected: The nine items for this construct loaded on two principal components. Therefore, we used two factors in subsequent analysis. The first factor, *Metrics_Basic* (alpha = 0.79) refers to the frequency with which simple cost, effort, and schedule metrics are collected in the organization. The second factor, *Metrics_Advanced* (alpha = 0.85) pertains to more sophisticated metrics such as quality measures, customer satisfaction, and requirements stability metrics.
2. Data Collection Procedures: (*Data_Collection*) alpha = 0.85, three items.

TABLE 2c
Questionnaire Items and Factor Loadings (cont.)

Item	Factor Loadings
Communication and Feedback	
How well do the following statements describe your organization?	
Communication	
a) Measurement status and results are provided to management on a regular basis	0.83
b) Status and results are provided to teams and technical staff on a regular basis	0.85
c) Results are communicated in time to be used for their intended purposes	0.79
d) Status and results are communicated through formal means, i.e. progress reports, memos, written reports etc.	0.77
Feedback	
a) Measurement reports are reviewed with all affected parties before they are reported more widely in the organization	0.83
b) Measurement data and analysis results are generally understandable and easily interpretable	0.83
Availability of Automated Tools (Auto-Tools)	
How much automated support is available for software measurement related activities in your organization?	
a) Data collection, e.g. on-line forms with tickler reminders, time stamped activity logs, static or dynamic analyses of call graphs etc.	0.84
b) Data management, e.g.database packages, tools for data integrity, Verification and validation	0.86
c) Data analysis and report preparation, e.g. speadsheets, statistical, graphing and report presentation packages	0.80
Alignment of Management in Goal-Setting (Align_other)	
How would you characterize the involvement of various potential stakeholders in setting goals and deciding on plans of action for measurement in your organization?	
a) Senior enterprise and organization level managers	0.68
b) Project level managers	0.73
c) Individual engineers, programmers or other practitioners	0.75
d) Business support units, eg. finance, marketing	0.60
Alignment of Measurement and SEPG in Goal-Setting (Align_QA)	
How would you characterize the involvement of various potential stakeholders in setting goals and deciding on plans of action for measurement in your organization?	
a) Technical support units	0.84
b) Measurement specialists	0.84

TABLE 2d
Questionnaire Items and Factor Loadings (cont.)

Item	Factor Loadings
Management Support (Mgmt)	
a) Management regularly monitors the progress of software measurement activities	0.92
b) Management clearly demonstrates commitment to measurement	0.93
Resource Sufficiency (Res_Suff)	
a) Is sufficient funding available for measurement-related activities in your software organization?	0.73
b) How much access does your organization have to library resources on software measurement and quantitative analysis?	0.70
c) Are qualified, well-prepared people available to work on software measurement in your organization?	0.82
Institutional Beliefs (Institution)	
How well do the following organizations describe your organization?	
a) The effort required for people to submit data is often considered to onerous or burdensome (reverse scored)	0.75
b) Software measurement and data analysis are an integral part of the way we normally do business	0.61
c) The way software measurement data are collected and used is often considered inappropriate by the people who must provide the necessary information (reverse scored)	0.77
d) There is resistance to doing measurement here, i.e. people think of it as unnecessary, extra work, unfair etc. (reverse scored)	0.82
Single Item Variables	
Training	
How would you best characterize the measurement related training in your organization?	
1 – *Excellent* to 5 - *Poor*	
Cmm_level	
To the best of your knowledge, what is the software process maturity level of your organization?	
1 – CMM Level 1	
2 – Close to Level 2	
3 – CMM Level 2	
4 – Close to Level 3	
5 – CMM Level 3	
6 – Higher	
7 – Don't Know	
SoftDev	
Approximately how much of your organization's business is devoted to software development (in percentages)?	

3. Quality of Metrics Collected: (*Metrics_Quality*) alpha = 0.70, two items.
4. Kinds of Analyses Done: (*Analyses*) alpha = 0.82, four items.
5. Training: (*Training*) one item.
6. Availability of Automated Tools: (*Auto_Tools*) alpha = 0.79, three items.
7. Communication and Feedback: Exploratory factor analysis was run on the six communication and feedback variables and the items loaded on two factors. The first factor, called *Comm*, (alpha = 0.83), measures the extent to which different communication patterns exist in the organization to provide metrics analysis and results. The second construct, *Feedback* (alpha = 0.58), measures the quality of feedback provided to the different people in the organization. This construct is the only one in our analysis that does not display adequately high reliability although the factor loadings on an underlying factor are high.

Organizational Factors:

1. Alignment of Personnel to Metrics Goals: As shown in Table 2, six questionnaire items were used to measure this construct. During principal components analysis, they loaded on two factors. Therefore, we used the composite scores on these two factors in subsequent analysis. The first factor, *Align_QA* (alpha = 0.85) measures the extent to which the measurement specialists and SEPG personnel contributed toward setting the goals for the metrics programs. The second factor, *Align_Other* (alpha = 0.71) measures the extent to which upper management and project managers were involved in setting goals.
2. Resource Sufficiency: (*Res_Suff*) alpha = 0.70, three items.
3. Management Support: (*Mgmt*) alpha = 0.83, two items.
4. Institutional Beliefs: (*Institution*) alpha = 0.73, four items.
5. Maturity Level: In order to measure the maturity level of the organizations, we use self-reported CMM level. This is not a perfect measure since there is a possibility of response bias. Moreover, it is possible for an organization's management to believe that they are at a certain maturity level without any external ratification. Therefore, although we use this variable as a measure of maturity, care should be taken in drawing inferences from the results. The responses on the seven categories in the questionnaire were not well distributed. Therefore, we dichotomized the variable using the following rule:

Level 2 and above were coded 1 while organizations below Level 2 were coded 0. The final variable used in the analyses, *Cmm_level*, is thus a binary variable.

6. Percentage Software Development: Our sample includes defense contractors, commercial vendors, and in-house MIS departments, not all of whom develop software as their primary business. Therefore, we hypothesize that the level of returns from metrics programs should be proportional to the percentage of effort dedicated by the organization to software development. In other words, organizations that do not develop software, as their primary activity may not gain as much benefit from a metrics program as an organization that spends most of its resources in software development. To test this hypothesis, we include a variable that measures the percentage of the organization's business devoted to development as opposed to maintenance or acquisition. This variable is called *Soft_dev*.

5 Analysis and Results

5.1 Specification of the Model

Given our theoretical model in Fig. 1, we use a simultaneous equation specification for our analysis. The simultaneous equation specification can estimate two equations, taking into consideration the relationship between the two dependent variables. In the presence of simultaneity in the two equations, estimation of the two equations separately gives biased and inefficient estimates of the regression coefficients [16]. The two equations to be estimated are given below.

$$\begin{aligned} Use\_Mgmt = {} & \alpha_0 + \alpha_1 Perf + \alpha_2 Metrics\_Basic \\ & + \alpha_3 Metrics\_Advanced + \alpha_4 Data\_Collection \\ & + \alpha_5 Metrics\_Quality + \alpha_6 Analyses \\ & + \alpha_7 Training + \alpha_8 Auto\_Tools + \alpha_9 Comm \\ & + \alpha_{10} Feedback + \varepsilon_1, \end{aligned}$$

$$\begin{aligned} Perf = {} & \beta_0 + \beta_1 Use\_Mgmt + \beta_2 Align\_QA + \beta_3 Align\_Other \\ & + \beta_4 Res\_Suff + \beta_5 Mgmt + \beta_6 Institution \\ & + \beta_7 Cmm\_level + \beta_8 Soft\_dev + \varepsilon_2. \end{aligned}$$

The error terms in the two equations are given by ε_1 and ε_2, respectively. Since the model specification is simultaneous, we use instrumental variables to estimate the appropriate regression coefficients for each of the individual independent variables. The instrumental variables used in our analysis were all the independent variables, i.e., the variables for technical and organizational factors, except the endogenous dependent variables. Since the data across the two equations are from the same organization and from the same respondent, we can assume that the two errors ε_1 and ε_2 are correlated and use this information to get results that are asymptotically efficient [16]. Therefore, we use iterated three-stage least squares to obtain estimates of the regression coefficients.[2] The estimates of the coefficients and their standard errors are shown in Table 3 along with model fit statistics.

TABLE 3
Iterated 3SLS Results $N = 214$

Equation 1: Dependent Variable = Use in Decision Making			
Variable	Coefficient	Std Error	Prob> |t|
Constant (α0)	- 0.327	0.209	0.118
Perf (α1)	0.649	0.123	0.0001
Metrics_Basic (α2)	0.098	0.037	0.009
Metrics_Advanced (α3)	0.138	0.043	0.001
Data_Collection (α4)	- 0.026	0.042	0.541
Metrics_Quality (α5)	- 0.033	0.046	0.470
Analyses (α6)	0.197	0.050	0.0001
Training (α7)	0.001	0.033	0.958
Automated_Tools (α8)	0.002	0.034	0.950
Communication (α9)	0.071	0.042	0.095
Feedback (α10)	-0.013	0.036	0.707
R-Square = 0.63		Adjusted R-square = 0.61	
Equation 2: Dependent Variable = Organizational Performance			
Variable	Coefficient	Std Error	Prob> |t|
Constant (β0)	1.404	0.406	0.0007
Use_Mgmt (β1)	0.316	0.129	0.015
Align_QA (β2)	0.074	0.036	0.039
Align_Other (β3)	0.275	0.072	0.0002
Res_Suff (β4)	0.049	0.043	0.262
Mgmt (β5)	0.081	0.041	0.049
Institution (β6)	0.185	0.075	0.014
Cmm_level (β7)	0.020	0.075	0.786
Soft_dev (β8)	0.002	0.001	0.050
R-Square = 0.54		Adjusted R-square = 0.52	

As shown by the results in Table 3, the simultaneous equation specification is borne out in the significant coefficients of the two dependent variables in the two equations. The individual coefficients in Table 3 measure the change in the dependent variable for a unit change in the independent variable. By assumption, the 3-stage procedure estimates a linear relationship between the independent variables and the dependent variable. In our discussion, we use an alpha of $p = 0.05$ to denote statistical significance. We start by discussing the results with respect to the cyclical relationship hypothesized between the two dependent variables.

5.2 Results and Discussion

In (1), the coefficient of Organizational Performance is positive and strongly significant ($p = 0.0001$). Similarly,

2. Residuals from three-stage least squares can be used iteratively to obtain newer estimates of the variance-covariance matrix between the two error terms. This estimation is called iterated 3SLS and has the same asymptotic properties of 3SLS [15]. In our case, the results from 3SLS and iterated 3SLS are similar.

in (2), the coefficient of Use in Decision-making is positive and significant, thereby indicating that the relationship between the two dependent variables is interdependent. An increase in the use of metrics information in decision making leads to higher organizational performance and similarly, an increase in organizational performance is associated with an increase in the use of metrics information in decision making. This supports our hypothesis of a cyclical relationship between the two dependent variables. We discuss the other variables one equation at a time.

5.2.1 Use in Decision Making

As expected, the frequency with which actual metrics collected has a positive and significant influence on their use in decision making. The presence of a systematic process of collecting metrics across the organization increases the use of this information in decision-making ($\alpha_2 = 0.098$, $\alpha_3 = 0.138$). The results also show that the marginal benefit from increased use of the variables constituting *Metrics_Advanced* is higher than the marginal benefit from *Metrics_Basic*, all else being equal. However, note that the basic metrics need to be in place for conducting metrics operations as captured in *Metrics_Advanced*. According to the Capability Maturity Model (CMM), organizations should address the low-level process areas before tackling the higher-level processes in their movement up the maturity ladder [22]. The metrics collected in *Metrics_Basic* are typically found in the Level2 specification of the CMM, whereas the metrics in *Metrics_Advanced* are typically associated with higher-level processes. Theoretically, the organization should experience greater returns from processes that are higher up in the maturity ladder and we see some support for that proposition here. We believe that the collection of metrics in *Metrics_Basic* should normally precede the collection of metrics in *Metrics_Advanced* and having gained some benefit from *Metrics_Basic*, the organization gains second order benefits from investment in *Metrics_Advanced*.

The Analyses variable is positive and strongly significant ($p = 0.0001$), thereby indicating that the rigor and usefulness of analyses done on the raw metrics data increases the use of metrics in decision-making. This seems intuitive since more sophisticated analyses cannot only help in decisions of the ordinary variety but can also help in making more difficult strategic decisions. For example, systematic and periodic analyses of metrics data can provide insights for software managers to identify gaps in current employee skills and build appropriate skill base in line with the business plans of the organization. For example, if an organization plans to focus on financial applications, a manager may be able to identify the right skill set for the new project from the metrics knowledge base of the past projects in this domain. In addition, metrics implementation across the organization may also enable managers to identify under performers and help them through customized training programs. Similarly, managers can spot high achievers to assign them to new responsibilities.

As noted earlier, we expected that better communication processes for the dissemination of metrics analysis should be associated with an increase in its use in decision making. Although the coefficient of Communication (*Comm*) variables in our analysis is in the expected direction ($\alpha 9 = 0.07$), it is not statistically significant at the $p = 0.05$ level. Hence, readers should interpret this result with care. The literature has presented some anecdotal evidence highlighting the importance of timely and clear communication of metrics results to the concerned parties and we see some support of that factor in our analysis [12], [19].

The other variables in this equation, *Data_Collection*, *Metrics_Quality*, and *Feedback* are not significant and, therefore, we cannot make any statements regarding their influence on the use of metrics on decision making. Further work is required to examine these variables in more detail and tease out their influences, if any, on metrics program success.

5.2.2 Impact on Organizational Performance

In (2), the alignment variables *Align_Other* and *Align_QA* are both positive and significant. Prior literature presents anecdotal evidence in support of the importance of incentive alignment with metrics usage and stresses the need to increase the involvement of even nonmeasurement related people in setting the goals of a metrics program [23], [7], [19]. Our results support this observation, indicating that involvement of upper management and project managers as well as metrics personnel increases organizational performance, all else being equal. The Goal-Question-Metric paradigm mandates setting goals for the metrics programs and stresses the need to involve all stakeholders in setting the goals before actual implementation is carried out [4]. The significance of the alignment variables in our model provides evidence in support of the GQM viewpoint.

The management support coefficient is also positive and significant ($p = 0.05$). This can be an extremely important factor in determining the success of a metrics program, and is clearly associated with organizational performance. In the innovation diffusion literature cited earlier, management influence plays an important role in deciding the rate of diffusion [15]. Management support is also important during the initial period of the metrics initiative when the results from the investment in metrics are not clear. It is usually during this period that most metrics programs terminate. Management support can not only ensure the existence of the program during this phase but can also enforce the use of metrics information in decision making until such time as tangible benefits are not visible.

The institutional belief factor is positively associated with organizational performance and is statistically significant. Institutionalization is a strong force that acts on most organizations. The need to conform to institutional forces is strong in organizations and, therefore, a prevalent belief in the values of metrics programs in an organization will enhance the use of metrics in decision making. Our results show that an organization that believes that metrics programs and information is beneficial has higher organizational performance than an organization without such a set of beliefs, all else being equal.

The *Soft_dev* variable, measuring the share of a company's resources being used in software development, is significantly and positively related to impact on organizational performance. Survey respondents from companies with higher percentages of software development as their primary activities are more likely to report that the use of software measurement has led to performance improvements in their organizations. We included this variable as a control measure and the result seems intuitive. Software development is generally considered riskier than maintenance or acquisition and the use of any processes or practices that help in decision making should have a greater benefit for development businesses than for maintenance or acquisition.

The resource sufficiency variable was not significant and so we cannot reach any conclusions about this variable's importance with respect to success. It is our belief that the results of our analysis should be used in assessing the relative effects of the various factors in our model rather than measuring their absolute effects on metrics programs. The other variable, *Cmm_level*, is also not statistically significant. As mentioned before, this variable is self-reported and there is a possibility of high response bias and measurement errors. Moreover, we had to dichotomize the variable due to its uneven distribution and therefore might have lost some information. Further research is required to test these two constructs in a different way to validate our findings.

6 Conclusion

In summary, we have addressed the important issue of analyzing the success of metrics programs in this paper. As previous literature has pointed out, the introduction and successful management of a metrics program has a high potential payoff but is fraught with problems. As noted earlier, findings from prior literature were either case based or anecdotal. To our knowledge, this is one of the first attempts to assess the success of software metrics programs in an empirical study with data from metrics experience in multiple organizations. Thus, one of the contributions of this paper is the introduction and empirical validation of some measures for the constructs discussed in past anecdotal research. Another significant contribution of this paper is that we use methodological and theoretical frameworks from other research streams such as innovation diffusion to adopt a two-staged process in studying metrics program success in software.

We believe that the results of our study can provide some useful insights for both practitioners and researchers. For practitioners, our results highlight the need to cultivate a culture within the organization to promote metrics usage to ensure eventual success of the metrics program. Our two-stage model identifies that software managers first need to focus on the technical factors and provide incentives for software developers to use the metrics information in daily operations. This will also facilitate the on-going collection of metrics from new projects. Once the metric data is used at the project level, managers can then focus on the organizational performance improvement in the second stage. Note that use of metrics data may be just one of the factors contributing to the improvement in organizational performance. Managers need to assess the various organizational factors that are identified as important in our study. For example, prompt usage of metrics-based analyses and feedback on the same and alignment of management incentives in metrics-based goal setting provides the base for reaping organizational performance improvement. Our study also highlights the importance of stressing metrics usage on a continuous basis to facilitate improvement in organizational performance in an on going basis.

Our research design and sampling inherently present some shortcomings. Note that the data used in this paper is one-time survey data; hence, we cannot carry out time series analysis at a greater level of detail. The interrelationships between the various success factors and their effect on metrics program success could be better studied in greater depth using more detailed, time-dependent data. The survey methodology also introduces some subjectivity in the responses and is open to recollection and response bias. Theoretically, a large sample size can mitigate these problems. From a behavioral science perspective, our sample size (214) and response rate (55 percent) are considered good. Moreover, our survey questionnaire was pretested thoroughly. Therefore, we believe that the weaknesses of the survey methodology have been mitigated to a great extent in our research. However, our model needs to be further validated in other data samples to assess the potential sample and response bias. Yet, another limitation with survey data is the potential for measurement errors due to the subjective nature of the variables measured. A more rigorous methodology to better control such measurement errors may be appropriate for future studies.

Note that this paper draws on anecdotal evidence, which is generally influenced by experience and to an extent, ideology rather than extant theory. Although this is one of the initial efforts to present some empirical support based on metrics experiences from various organizations, our model does not discuss the theory behind potential interaction across various success factors. Since implementing software metrics can be viewed as a major change within software organization, a direction for future research may be to formulate a model well grounded in organization innovation and change management theories that captures the complex interactions across various metrics success factors.

Acknowledgments

Anandasivam Gopal was supported by the William Larimer Mellon Fellowship at the Graduate School of Industrial Adminstration, and a Research Assistantship at the Software Engineering Institute. The Software Engineering Institute (SEI) is a federally funded research and development center sponsored by the US Department of Defense and operated by Carnegie Mellon University.

References

[1] I. Aaen, A. Siltanen, C. Sorensen, and V.-P. Tahvanainen, "A Tale of Two Countries: CASE Experiences and Expectations," *The Impact of Computer Supported Technologies on Information Systems Development, IFIP Transactions,* K.E. Kendall, K. Lyytinen, and J. DeGross, eds., pp. 61-91, 1992.

[2] I. Aaen, "Problems in CASE Introduction: Experiences from User Organizations," *Information and Software Technology,* vol. 36, pp. 643-654, Nov. 1994.

[3] J.V. Baldridge and R.A. Burnham, "Organizational Innovation: Individual, Organizational, and Environmental Impacts," *Administrative Science Quarterly,* vol. 20, pp. 165-176, June 1975.

[4] V.R. Basili and D.M. Weiss, "A Methodology for Collecting Valid Software Engineering Data," *IEEE Trans. Software Eng.,* vol. 10, pp. 728-738, 1984.

[5] J.C. Brancheau and J.C. Wetherbe, "The Adoption of Spreadsheet Software: Testing Innovation Diffusion Theory in the Context of End-User Computing," *Information Systems Research,* vol. 1, pp. 115-143, June 1990.

[6] L.C. Briand, C.M. Differding, and H.D. Rombach, "Practical Guidelines for Measurement-Based Process Improvement," *Software Process—Improvement and Practice,* vol. 2, pp. 253-280, 1996.

[7] M. Daskalantonakis, "A Practical View of Software Measurement and Implementation Experiences Within Motorola," *IEEE Trans. Software Eng.,* vol. 18, pp. 998-1010, Nov. 1992.

[8] S. Fenick, "Implementing Management Metrics: An Army Program," *IEEE Software,* pp. 65-72, Mar. 1990.

[9] P. Finlay and A.C. Mitchell, "Perceptions of the Benefits from the Introduction of CASE: An Empirical Study," *MIS Quarterly,* vol. 18, pp. 353-370, Dec. 1994.

[10] T. Hall and N. Fenton, "Implementing Effective Software Metrics Programs," *IEEE Software,* pp. 55-65, Mar./Apr. 1997.

[11] J. Iivari, "Why are CASE Tools Not Used?" *Comm. ACM,* vol. 39, pp. 94-103, Oct. 1996.

[12] J. Iversen and L. Mathiassen, "Lessons from Implementing a Software Metrics Program," *Proc. Hawaii Int'l Conf. System Sciences (HICSS-33),* 2000.

[13] R. Jeffery and M. Berry, "A Framework for Evaluation and Prediction of Metrics Programs Success," *Proc. First Int'l Software Metrics Symp.,* pp. 28-39, 1993.

[14] C.F. Kemerer, "How the Learning Curve Affects CASE Tool Adoption," *IEEE Software,* vol. 9, pp. 23-28, 1992.

[15] J. Kimberley and M.J. Evanisko, "Organizational Innovation: The Influence of Individual, Organizational, and Contextual Factors on Hospital Adoption of Technological and Administrative Innovations," *Academy of Management J.,* vol. 24, pp. 689-713, Dec. 1981.

[16] J. Kmenta, *Elements of Econometrics.* New York: MacMillan Publishing, 1989.

[17] S.K. Majumdar and S. Venkataraman, "New Technology Adoption in US Telecommunications: The Role of Competitive Pressures and Firm-Level Inducements," *Research Policy,* vol. 22, pp. 521-536, 1993.

[18] S.K. Majumdar, "Does New Technology Adoption Pay? Electronic Switching Patterns and Firm-Level Performance in US Telecommunications," *Research Policy,* vol. 24, pp. 803-822, 1995.

[19] J. Mayes, "Achieving Business Objectives III: A Real-World Software Process Improvement Implementation," *IT Metrics Strategies,* vol. VI, no. 8, Aug. 2000.

[20] J.W. Meyer and B. Rowan, "Institutionalized Organizations: Formal Structure as Myth and Ceremony," *Am. J. Sociology,* vol. 83, pp. 340-363, 1983.

[21] R.J. Offen and R. Jeffery, "Establishing Software Measurement Programs," *IEEE Software,* pp. 45-53, Mar./Apr. 1997.

[22] M.C. Paulk, B. Curtis, M.B. Chrissis, and C.V. Weber, "Capability Maturity Model for Software, Version 1.1," Technical Report SEI-93-TR-24, Software Eng. Inst., Carnegie Mellon Univ., 1993.

[23] S.L. Pfleeger, "Lessons Learned in Building a Corporate Metrics Program," *IEEE Software,* pp. 67-74, May 1993.

[24] G. Premkumar, K. Ramamurthy, and S. Nilakanta, "Implementation of Electronic Data Interchange: An Innovation Diffusion Perspective," *J. Management Information Systems,* vol. 11, pp. 157-186, Aug. 1994.

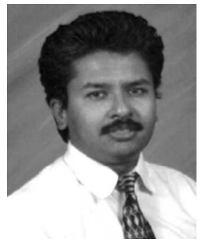
[25] K. Ramamurthy and G. Premkumar, "Determinants and Outcomes of Electronic Data Interchange Diffusion," *IEEE Trans. Eng. Management,* vol. 42, pp. 332-351, Nov. 1995.

[26] L. Rosenberg and L. Hyatt, "Developing a Successful Metrics Program," *Proc. Eighth Ann. Software Technology Conf.,* Apr. 1996.

[27] R. Rosenthal and R.L. Rosnow, *Essentials of Behavioral Research: Methods and Data Analysis,* second ed. McGraw-Hill, 1991.

[28] C. Seddio, "Integrating Test Metrics within a Software Engineering Measurement Program at Eastman Kodak Company: A Follow-Up Case Study," *J. Systems & Software,* vol. 20, pp. 227-235, 1993.

Anandasivam Gopal received an undergraduate degree from the Birla Institute of Technology and Science, Pilani, India, and the master's degree from the University of North Carolina, Chapel Hill. He received the PhD degree in information systems from the Graduate School of Industrial Administration, Carnegie Mellon University, Pennsylvania, in 2000. He is an e-business engineering associate at PRTM, a consulting firm specializing in consulting for the high-tech sector. His interests lie in the economics of software development and empirical software engineering, more specifically in software contracting and metrics programs in software development organizations. He has also worked as a research associate at the Software Engineering Institute, affiliated to Carnegie Mellon University.



M.S. Krishnan received the PhD degree in information systems from the Graduate School of Industrial Administration, Carnegie Mellon University, Pennsylvania, in 1996. He is a Mary and Mike Hallman e-Business Fellow, director of eLab, and associate professor of Computer Information Systems at the University of Michigan Business School. Dr. Krishnan was awarded the ICIS Best Dissertation Prize for his Doctoral Thesis on "Cost and Quality Considerations in Software Product Management." His research interest includes modeling of issues related to customer satisfaction, quality and productivity in information systems, business value of information technology, return on investments in software process improvements, software engineering economics, metrics, and measures for quality and customer satisfaction for products in software and information technology industries. His research articles have appeared in several journals, including *Management Science*, *Information Technology and People*, *Harvard Business Review*, *IEEE Transactions on Software Engineering*, *Decision Support Systems*, *Information Week*, and *Communications of the ACM*.

Tridas Mukhopadhyay is Deloitte Consulting Professor of e-Business at Carnegie Mellon University (CMU), Pennsylvania. He is also the director of the Institute for Electronic Commerce and MSEC (Master of Science in Electronic Commerce) program at CMU. He received the PhD degree in computer and information systems from the University of Michigan in 1987. His research interests include business-to-business commerce, Internet use at home, business value of information technology, and software development productivity. His recent papers on e-business examine loyalty on the Internet, and competition between Internet search engines. In addition, Professor Mukhopadhyay has worked with several organizations, such as General Motors, Ford, Chrysler, PPG, Port of Pittsburgh, Texas Instruments, IBM, Diamond Technology Partners, LTV Steel, and governmental agencies including the United States Post Office and the Pennsylvania Turnpike. He has published more than sixty papers. His research appears in *Information Systems Research*, *Communication of the ACM*, *Journal of Manufacturing and Operations Management*, *MIS Quarterly*, *Omega*, *IEEE Transactions on Software Engineering*, *Journal of Operations Management*, *Accounting Review*, *Management Science*, *Journal of Management Information Systems*, *Decision Support Systems*, *Journal of Experimental and Theoretical Artificial Intelligence*, *Journal of Organizational Computing*, *International Journal of Electronic Commerce*, *American Psychologist*, and other publications.

Dennis R. Goldenson is a senior member of the technical staff in the Software Engineering Measurement and Analysis group at the Software Engineering Institute in Pittsburgh, Pennsylvania. His work focuses on the use of measurement and analysis in software engineering, the improvement of process appraisal methods and models, and the impact and transition of software process improvement and other software engineering practices. Related interests are in tools to support collaborative processes, survey research, and experimental design. Dr. Goldenson is a principal author of the Measurement and Analysis Process Area for CMMI^{sm} and currently serves as colead of test and evaluation for the project. Additionally, he is the international trials coordinator for empirical methods of the SPICE project in support of ISO/IEC 15504. Other recent work has included requirements elicitation for training on the acquisition of software intensive systems and a study of the practices of high maturity organizations.

▷ **For more information on this or any computing topic, please visit our Digital Library at** http://computer.org/publications/dlib.