A CORRELATIONAL STUDY OF THE SEI'S CAPABILITY MATURITY MODEL AND SOFTWARE DEVELOPMENT PERFORMANCE IN DOD CONTRACTS

THESIS

Robert M. Flowe, Captain, USAF
James B. Thordahl, Captain, USAF

AFIT/GSS/LAR/94D-2

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY
AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

AFIT/GSS/LAR/94D-2

A CORRELATIONAL STUDY OF THE SEI'S CAPABILITY MATURITY MODEL AND SOFTWARE DEVELOPMENT PERFORMANCE IN DOD CONTRACTS

THESIS

Robert M. Flowe, Captain, USAF
James B. Thordahl, Captain, USAF

AFIT/GSS/LAR/94D-2

AFIT/GSS/LAR/94D-2

A CORRELATIONAL STUDY
OF THE SEI'S
CAPABILITY MATURITY MODEL
AND SOFTWARE DEVELOPMENT PERFORMANCE
IN DOD CONTRACTS

THESIS

Presented to the Faculty of the Graduate School of Logistics and
Acquisition Management of The Air Force Institute of Technology
AETC
In Partial Fulfillment of the Requirements for the Degree of
Master of Science In Software Systems Management

Robert M. Flowe, B.S.
Captain, USAF

James B. Thordahl, B.S.
Captain, USAF

December 1994

Approved for public release; distribution unlimited

Preface

We chose to investigate the correlation between software process maturity, as measured by the Capability Maturity Model (CMM), and software project success in response to a proposed policy requiring bidders on all Air Force software contracts to be rated CMM Level 3. This policy, promulgated by the Office of the Secretary of the Air Force (SAF/AQK), was proposed despite the absence of substantial quantitative data supporting the presumption that higher rated contractors are necessarily more successful in terms of cost and schedule performance than lower rated contractors. Our goal was to explore the nature of the correlation in an empirical manner, to provide a degree of rigor in the analysis of the presumed correlation between rating and performance. We hope the results of our research further the understanding of the use of the CMM as a means to assess the general likelihood of a contractor's software project success, and provide the software acquisition manager with another tool in the ongoing battle against late, over-budget software.

We would like to thank our thesis advisors: Dr. Freda Stohrer for her technical writing expertise and ongoing enthusiasm, and Lt Col Pat Lawlis for keeping us on track and providing a real-world software perspective. We would also like to thank Professor Dan Reynolds for his guidance in proper use of nonparametric statistical analysis methods, and for sharing with us his commitment to knowledge. Thanks also to Dr. David Christensen for his guidance in the area of cost analysis and the C/SCSC.

For their unstinting support and detailed knowledge, we thank the personnel of the ESC and ASC cost libraries, in particular, Mary Dutra and Sandra McCardle. Without their support, we would have been lost in a sea of data.

Though they must remain anonymous, we extend special thanks to the many organizations that participated in our research. These busy people gave of their time and knowledge to provide us with the critical information necessary to derive understanding from our data. Many of them encouraged us by expressing interest in our study, reinforcing our commitment to our audience--the project managers of the Air Force.

For inspiring us to undertake this research, and providing us with valuable guidance, we thank Captain Brian Hermann, and Captain Raymond Lewis, Jr., whose research provided the impetus for our thesis.

Finally, we thank our wives and families. Without their understanding and patience, this research would not have been possible.

Captain Robert M. Flowe

Captain James B. Thordahl

Table of Contents

Page

Page

Page

Page

Page

List of Figures

Page

Page

Page

List of Tables

Page

List of Acronyms

ACWP	Actual Cost of Work Performed
AFB	Air Force Base
AFMC	Air Force Materiel Command
AOV	Analysis of Variance
ASC	Aeronautical Systems Center
BAC	Budget at Complete; Budget at Completion
BCWP	Budgeted Cost of Work Performed
BCWS	Budgeted Cost of Work Scheduled
CMM	Capability Maturity Model
CPI	Cost Performance Index
CPFF	Cost Plus Fixed Fee
CPIF	Cost Plus Incentive Fee
CPR	Cost Performance Report
CWBS	Contract Work Breakdown Structure
C/SCSC	Cost/Schedule Control Systems Criteria
DoD	Department of Defense
DSI	Delivered Source Instructions
ECP	Engineering Change Proposal
ESC	Electronic Systems Center
FP	Fixed Price [Contract]
HOL	High Order Language
KSLOC	Thousand Source Lines of Code
K-W	Kruskal-Wallis
LOC	Lines of Code
LRE	Latest Revised Estimate
RDT&E	Research, Development, Test and Evaluation
REVIC	Revised, Enhanced Version of Intermediate COCOMO
RFP	Request for Proposal
SCE	Software Capability Evaluation
SDCCR	Software Development Capability/Capacity Review
SDCE	Software Development Capability Evaluation
SEI	Software Engineering Institute
SLOC	Source Lines of Code
SPA	Software Process Assessment
SPI	Schedule Performance Index
SPO	Systems Program Office
WBS	Work Breakdown Structure

AFIT/GSS/LAR/94D-2

Abstract

The Software Engineering Institute's (SEI's) Capability Maturity Model (CMM) is widely used to measure an organization's software development process maturity. The Department of Defense (DoD) has adopted this model with the belief that a more mature software development process will result in a more successful software project. Although there is a growing body of anecdotal evidence supporting this presumed correlation, there is currently no empirical evidence. Thus, the goal of our research was to determine the nature of the correlation, if any, between software process maturity and software project success, where process maturity is based on a CMM rating and success is based on the parameters of cost and schedule. To investigate this correlation we identified software unique projects, obtained CMM rating information on the contractor, collected cost and schedule data from a time frame representative of the rating, and interviewed project personnel to collect project context information. Using plots of cost and schedule performance versus rating level and nonparametric statistical techniques we found that, within our dataset, a correlation does exist between software development process maturity and project performance. The nature of this correlation appears to be improved cost and schedule performance with higher software process maturity.

A CORRELATIONAL STUDY OF THE SEI'S CAPABILITY MATURITY MODEL AND SOFTWARE DEVELOPMENT PERFORMANCE IN DOD CONTRACTS

1. Introduction

1.1 General Issue

The Department of Defense (DoD) is profoundly dependent upon computer software--a critical element of virtually every weapon system the DoD operates. DoD's reliance on software-intensive systems is increasing dramatically as the DoD tries to maximize the effectiveness of systems procured with the dwindling acquisition budget. Unfortunately, the trend in software-intensive procurements has been late, over-budget systems which fall short of customers' requirements. No single root cause has been identified for these programmatic failures, though resolving this dilemma is a high priority within the DoD.

In 1986, the DoD founded the Software Engineering Institute (SEI) as a center of excellence to address the problems that plague the procurement of DoD software. Key among its accomplishments is SEI's elaboration of a software development paradigm which holds that the quality of the software product is directly related to the maturity of the software development process. This concept is encapsulated in the SEI's Capability Maturity Model (CMM). The model characterizes an organization's process maturity based on several key characteristics, such as project management, configuration management, training, software quality assurance, and automation. This process of maturity assessment is formalized in standard protocols, which can be used by an organization to determine its own process maturity, or to assess the process maturity of another organization. The level of process maturity is expressed by a numerical rating, which runs from 1 (lowest process maturity) to 5 (highest process maturity). Organizations which have the highest process maturity are considered most likely to produce the highest-quality software.

The SEI's CMM has been widely accepted as a significant step toward solving the problems plaguing the development of DoD software (Mosemann, 1992:4). By applying the process maturity assessment protocols to a potential software developer, the government obtains an assessment of the developer's capability to produce quality software. Procurement risk is thus reduced, and the probability of obtaining the desired software within the constraints of schedule and budget are increased. The key assumption is that there is a significant positive correlation between SEI CMM rating and the success of the software development.

1.2 Specific Problem

Very little empirical research has been performed to establish a correlation between CMM rating and the success of software product development in terms of cost, schedule, and product quality (Hersh, 1993:12). However, there is a small but growing body of anecdotal evidence supporting the correlation between the SEI's CMM rating, process maturity, and some measures of software development success (Dion, 1993:28-32; Humphrey, Snyder, and Willis, 1991:11). These reports are generally self-reported assessments by DoD software development contractors. Although these reports show improvement in bottom-line issues such as return on investment, they do not generally address the issue of whether the DoD's interests are served by such process improvements. Notwithstanding these success stories, the lack of empirical evidence has prompted one critic to claim that "... it appears unlikely that such ratings have any meaningful correlation to the actual abilities of organizations to produce high quality software on time and within budget" (Bollinger, 1991:26).

1.3 Research Objectives

The purpose of our research is to establish the nature of the correlation, if any, between an organization's software development process maturity, as expressed by the SEI's CMM rating, and the success of the products the organization produces. For the purposes of our research, success is defined as the degree to which the project meets requirements, expressed in terms of cost performance, schedule performance, and quality parameters.

1.4 Scope/Limitations

We chose a research methodology which provided the greatest opportunity for meaningful information within our time and resource constraints. This limited our study to those organizations which met the following criteria:

a. Developed software for the DoD.

b. Were rated according to the SEI CMM protocols.

c. Tracked cost, schedule, and quality data in a standard format.

d. Reported relevant data to the DoD.

The above operational constraints led us to select DoD contractor organizations which provide software for Air Force Systems Program Offices (SPOs), where the desired cost, schedule and quality data were reported to the Air Force as part of the terms of the contract and in accordance with standard methods.

1.5 Overview

Our research establishes the nature of the relationship, between the SEI's CMM rating and software development success in selected DoD contractor organizations. Our intent is to provide the DoD software acquisition manager with a valid basis for important software acquisition and management decisions relating to contractor software development capabilities. We do this by providing insight into how well CMM ratings correlate to successful product development. This correlation was investigated by collecting data on a number of DoD contractors who had been rated by the CMM protocols, and the software products these organizations have developed while this rating has been in effect. We focus on cost, schedule, and quality performance exhibited during these sample projects. We also explore possible moderating variables (such as project size, application type, language, maturity assessment method, etc.) which could influence the correlation. These moderators may substantially affect the outcome of the software development, and thus should be taken into account. Acknowledging the limitations of our research, we hope to begin a continuing effort to validate the SEI's model and thus provide the software acquisition manager a reliable indicator for software development success.

2. Literature Review

2.1 Introduction

The Department of Defense is concerned about software success (within the procurement realm), and software effectiveness (within the operational realm) given its dependence upon software-driven systems, and the dwindling monetary resources to acquire and maintain them. As a result, the DoD has placed increasing importance on the improvement of the software development process. In the following discussion, the concept of software development process maturity will be discussed. It is important to distinguish between the concept of the software development *process*, and the *maturity* of this process. The term *maturity* in this case refers to the degree of refinement and sophistication of the software development process.

First, the software development process will be introduced. Several software development process models used to characterize the software development process will be described as well. Next, the concept of software development process maturity will be introduced. The SEI, and the CMM it developed to characterize software development process maturity will then be described, followed by applications and limitations of the model. Measurement of software development process maturity is then outlined. Some of the current concepts surrounding the measurement of project success based on cost/schedule and quality measures will then be introduced, along with a discussion of using cost and schedule parameters to assess process maturity. Following this, some noteworthy case histories will be presented, wherein the application of the CMM paradigm resulted in significant process improvements. Limitations of the CMM will also be discussed. Finally, the latest information regarding correlations between process maturity and product effectiveness/success will be outlined

2.2 The Software Development Process

A distinction must be made between the concept of the *software development process* and the concept of *software development process maturity*. This section describes the concept of the software development process. Simply stated, a software process is the set of methods, tools, and practices used to produce software products (Humphrey, 1989:3). The software development process, in general, is the process by which abstract requirements, or user needs are transformed into concrete software products. The software development process can be characterized by one of several popular models.

2.3 Software Development Process Models

Software development process models are symbolic constructs used to describe the process of transforming the abstract software requirements to concrete code. "Just as a software program defines a process that a computer must follow to achieve a result, software process models define the process a software engineer must follow" (Lai, 1993:16).

Several models have been developed to describe the software development process. Below are some of the more prominent software models currently in use.

2.3.1 The Waterfall Model

The Waterfall model, described by Royce in 1970 is still the best known and most widely used framework for the software development process (Humphrey, 1989:249). It has "...become the basis for most software acquisition standards in government and industry" (Boehm, 1988:63). The Waterfall model is featured in the DoD standard for software development, DoD-STD-2167A (DoD, 1988:10). The foundation of the waterfall is the sequential series of steps that translate abstract software requirements into a software product (Figure 2-1).

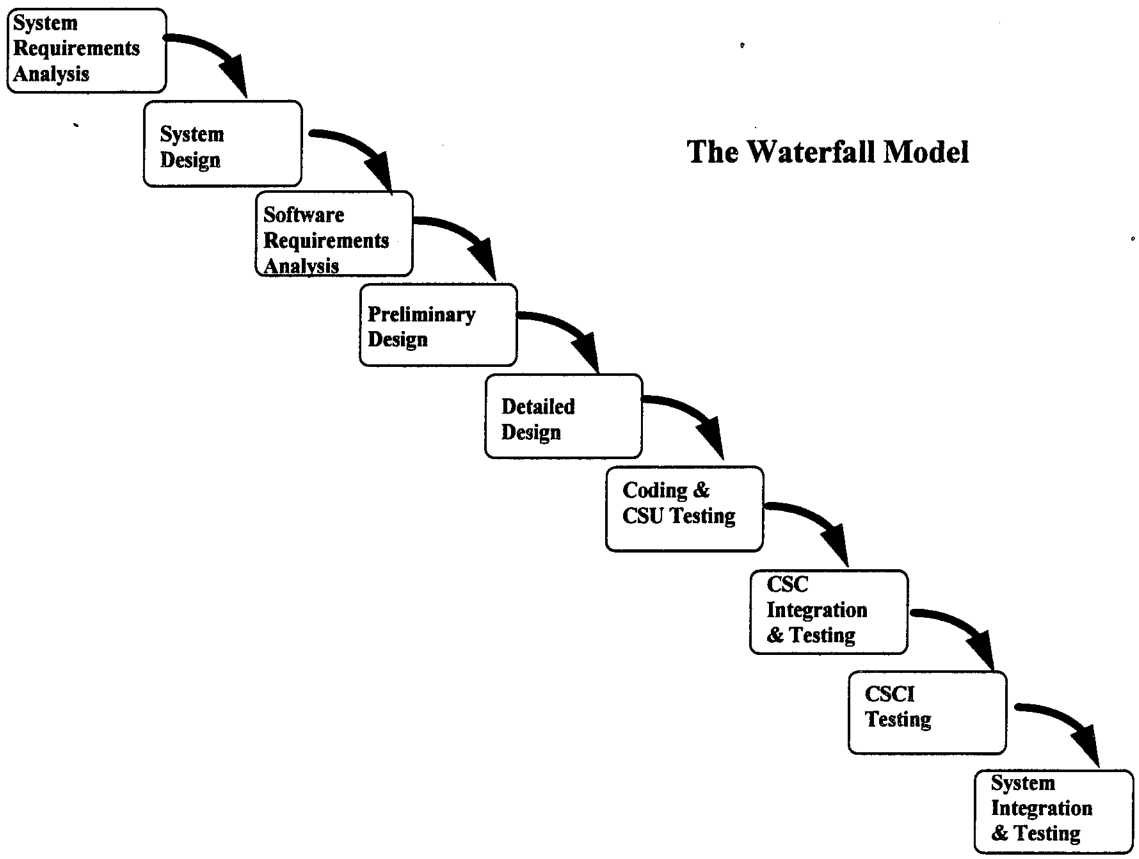

Figure 2- 1 The Waterfall Model

2.3.2 The Spiral Model

The Spiral model was proposed by Boehm in 1988, "...based on experience with various refinements of the Waterfall model as applied to large government software projects" (Boehm, 1988:64). The Spiral model (Figure 2-2) superimposes iterative risk identification and mitigation activities (such as risk analysis and prototyping) over the sequential software development steps of the Waterfall model.

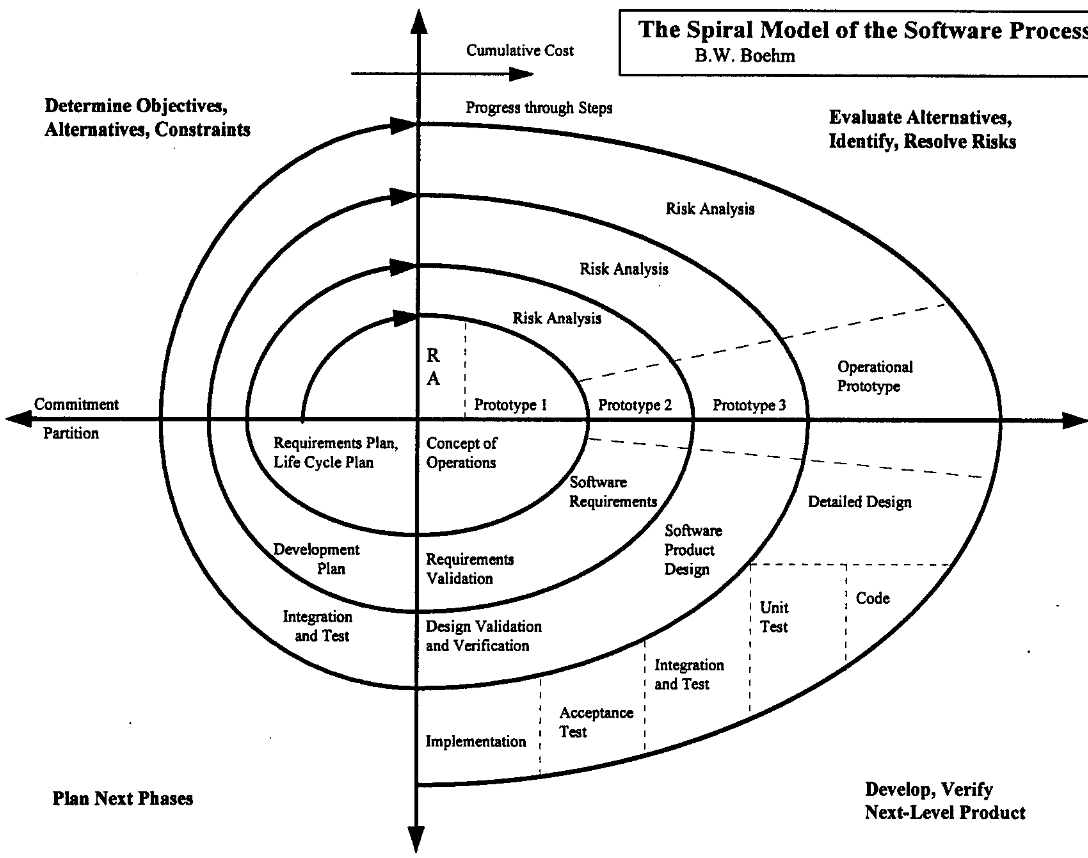

Figure 2- 2 The Spiral Model

2.3.3 Prototyping

A software prototype implements part of the presumed software requirements to learn more about actual requirements or about alternative designs that could satisfy the requirements (Davis, 1992:71). Prototyping can be used in several ways, either as an element of a software development process model (as with the Spiral model), or as a software development process model in its own right (as with Evolutionary and Operational Prototyping described below).

Throwaway Prototyping: In Throwaway Prototyping, the prototype is built as quickly as possible, typically without great attention to quality and standards, in order to get immediate feedback from the user. This helps to understand and elaborate ill-defined user requirements. This "quick and dirty" solution is not suitable for long-term or operational use, and thus must be "thrown away." The information gained from the user is then used to write the requirements specification for the follow-on system, which will subsequently be built in a more quality conscious manner.

Evolutionary Prototyping: Evolutionary Prototyping is a more quality-intensive approach wherein the prototype implements requirements that are well-understood, and confirmed. The prototype is used to identify unknown requirements. These requirements, when identified, are rolled into the software requirements specification. The system is then redesigned, recoded, and retested. The evolutionary development process is repeated indefinitely, with new prototypes, new requirements, and an evolving design.

Operational Prototyping: Operational Prototyping is a synthesis of evolutionary and throwaway prototyping, and is used in situations where neither approach alone would be appropriate. A typical operational prototype approach involves developing and fielding a quality-intensive system (conforms to standards--fully tested and documented) which incorporates basic, well-understood requirements, and constitutes the "baseline" design. At the user's site, prototype enhancements are generated in near-real-time, by prototypers interacting directly with users. These prototype enhancements are made in response to user input, either identifying new requirements, or refining poorly understood requirements. If these changes are found by the user to be effective, the requirements are incorporated into the design baseline. The prototype enhancement is thrown away when the new baseline is established. This process can continue indefinitely, with each successive enhancement adding new, user-driven capability.

2.4 Software Development Process Maturity

The concept of software development process *maturity* is used as a means of characterizing an organization's implementation of the software development process. Thus it is less dependent upon the software development process model (e.g. waterfall model, spiral model, etc.) than it is upon the particular organization with respect to how the organization implements and manages the model. For example, an organization may use a waterfall software development process model, the implementation of which may be either mature or immature, depending upon how the organization chooses to manage the process. The process *maturity* is a reflection of the organization itself, at a level more profound than the particular software development model to which the organization subscribes. A definition of software process maturity is "the extent to which a specific process is explicitly defined, managed, measured, controlled, and effective. Maturity implies a potential for growth in capability and indicates both the richness of an organization's software process and the consistency with which it is applied in projects throughout the organization" (Paulk, Curtis, Chrissis, and Weber, 1993:20).

The concept of process maturity evolved from the failure of software development process models to address the problems of late, over-budget, low quality software. Initially it was thought that by formalizing the software development process using these models, the problems of poor software would be resolved. It became clear that the formalization of software development models was not sufficient to create quality software. Consideration for how an organization uses a model is also necessary. To this end, the SEI was established by the DoD to introduce improved software development methods into general practice (Humphrey, Kitson, and Kasse, 1989:1).

2.5 The Capability Maturity Model

"In November 1986, the SEI, with assistance from Mitre Corp., began developing a process-maturity framework that would help developers improve their software process" (Paulk, Curtis, Chrissis, and Weber, 1993:18).

The framework developed by the SEI is based on two premises: "[First,] the process of producing and evolving software products can be defined, managed, measured, and progressively improved and [second] the quality of a software product is largely governed by the quality of the process used to create and maintain it" (Humphrey, Kitson, and Kasse, 1989:5). This process maturity framework is articulated in the SEI's CMM. The intent of the CMM is to provide a framework for characterizing a software development organization's process maturity. In his book Managing the Software

Process, Watts Humphrey details the five levels of process maturity contained in the CMM, the major points of which are summarized below.

Level 1--Initial - The initial process level could properly be called ad hoc, and is often chaotic. An initial-level organization often operates without formalized procedures, cost estimates, and project plans. Tools are neither well integrated with the process nor uniformly applied. Change control is lax, and there is little senior management exposure or understanding of the problems and issues. Organizations at the initial level can improve their performance by instituting basic project controls. The most important are project management, management oversight, quality assurance, and change control.

Level 2--Repeatable - The repeatable process provides control over the way the organization establishes its plans and commitments. This provides an improvement over the initial level, achieving a degree of statistical control through learning to make and meet their estimates and plans. The key actions required to advance from the Repeatable to the Defined process are to establish a process group, establish a development process architecture, and introduce a family of software engineering methods and technologies.

Level 3--Defined - The Defined process establishes the foundation for examining the process and deciding how to make improvements, thus opening the door for major and continuing progress. The qualitative nature of the Defined process, however, prevents the organization from measuring how much is accomplished, or how effective the process is. Key steps to advance from Defined to Managed are (1) to establish a minimum basic set of process measurements to identify the quality and cost parameters for each process step, (2) to establish a process database for cost and yield data, (3) to provide sufficient process resources to gather and maintain this process data, and to advise project members on its use, and (4) to assess the relative quality of each product and inform management where quality targets are not being met.

Level 4--Managed - the Managed process gathers the process data and makes informed decisions about the process. One of the biggest challenges to the Managed process is the cost of gathering data. Key steps for advancing from the Managed to the Optimizing process are (1) the support of automatic gathering of process data, to improve the accuracy and quality of the data, and (2) the use of process data to both analyze and modify the process to prevent problems.

Level 5--Optimizing - The transition from the Managed process to the Optimizing process represents a paradigm shift. Whereas the data collection and analysis for the Managed process was focused toward facilitating product improvements, with the Optimizing process, the data is collected and used to tune the process itself. With an Optimizing process, the organization now has the means to identify the weakest elements of the process and fix them. At this point, data is available to justify the application of technology to various critical tasks, and numerical evidence is available on the effectiveness with which the process has been applied (Humphrey, 1989:6-12).

According to Humphrey, the above levels were selected because they:

- represent the actual historical phases of a software organization's evolution,
- represent an achievable measure of improvement from the prior level,
- suggest interim improvement goals and progress measures, and
- identify a set of improvement priorities (Humphrey, 1989:5).

Thus, the characteristics of the above maturity levels enable the CMM to be used not only as a tool to assess an organization's current process maturity, but also to recommend avenues by which the process can be improved.

2.5.1 Process Maturity Measurement

The CMM consists of a hierarchical structure that allows an assessment team to evaluate an organization's process maturity. At the top are the five individual *maturity levels* describing how the organization is expected to function. With the exception of Level 1, the five maturity levels are decomposed into *key process areas* which indicate where an organization should focus to improve its process. For example, the key process areas at Level 2 focus on establishing basic project-management controls and include Requirements Management, Software Project Planning, Software Project Tracking and Oversight, Software Subcontract Management, Software Quality Assurance, and Software Configuration Management (Paulk, Curtis, Chrissis, and Weber, 1993:25).

Each key process area is composed of *key practices* that must be followed to satisfy the goals of the key process area. "Key practices describe the infrastructure and activities that contribute most to the effective implementation and institutionalization of the key process areas" (Paulk, Curtis, Chrissis, and Weber, 1993:26). Key practices can be viewed as the working definitions of the key process areas (Honour Werth, 1993:12).

At the heart of the assessment though, is the *maturity questionnaire*. The maturity questionnaire consists of questions that enable the assessment team to identify the presence or absence of key practices and determine whether the goals of the key process area are being satisfied. "Questions are not open-ended, but are intended to obtain a quantified result from following answers: yes, no, don't know, and not applicable" (Honour Werth, 1993:12). These initial responses serve as the basis for a more detailed open-ended question process between the assessment team and key members of the organization being evaluated.

The result of a process maturity assessment is the assignment of a maturity level rating. The assessment team uses the responses from both the personal interviews and the maturity questionnaire along with results of document reviews to determine if an organization is meeting the goals of specific key process areas. In order to attain a particular maturity level rating, such as Level 2, the organization must meet the goal of each Level 2 key process area identified in the CMM.

2.5.2 Applications of the CMM (SPA and SCE)

A process maturity model is merely of academic interest unless it can be meaningfully applied to real-world organizations. Thus, "the operational elaboration of the CMM is designed to support the many ways it will be used[,] four of which are[:]

- Assessment teams will use it to identify strengths and weaknesses in an organization.
- Evaluation teams will use it to identify the risks of selecting among contractors and to monitor contracts.

- Upper management will use it to understand activities necessary to launch a process-improvement program in their organization.
- Technical staff and process-improvement groups will use it as a guide to help them define and improve their organization's process" (Paulk, Curtis, Chrissis, and Weber, 1993:24).

The operational elaboration of the CMM is expressed by two distinct assessment methods, the first of which is the Software Process Assessment (SPA). A SPA is used by an organization to determine its own process maturity, gain insight into its development capability, and prioritize management actions for transition to the next maturity level. The second method is the Software Capability Evaluation (SCE). A SCE is an independent evaluation of an organization's process maturity to gain insight into its ability to produce domain specific software. A SCE is initiated and funded by the Government, and is used as a criterion during contract award (Besselman, Byrnes, Lin, Paulk and Puranik, 1993:6-7).

2.5.3 Limitations of the CMM and its Application

The CMM has been widely used as a framework for process assessment and improvement as well as a tool for bidder maturity assessment in Government procurement projects. However, the model is not without its limitations. As this model has been put into practice it has become evident to the SEI that a SPA and a SCE may not result in the same maturity level rating, because of differences in the motivation for the use of each rating method. As previously stated, a SPA is an internal assessment while a SCE is an external audit. This contrast in application leads to dissimilar approaches to certain aspects of the assessment or evaluation, including the selection of projects, the investigative methods used, and the level of familiarity with the development organization (Besselman, Byrnes, Lin, Paulk and Puranik, 1993:24).

Project Selection : A SPA is intended to characterize the organization's software development process maturity as a whole. As a result, projects are selected on their overall representativeness of the organization and may come from multiple software application domains. In contrast, a SCE is performed for the purpose of identifying an organization's software development capability with regards to a particular Government procurement. Project are thus selected for evaluation based on their similarity to the anticipated procurement.

Investigative Methods: A SPA utilizes one-on-one interviews and group discussions to determine the process maturity level of the organization under consideration. This is done in order to promote an organization's awareness of their maturity level while also encouraging an atmosphere of process improvement within the organization. A SCE, on the other hand, utilizes one-on-many interviews and relies heavily on document reviews to objectively determine the process maturity level of an organization. A SCE is not as much concerned with process improvement as it is the objective determination of an organization's software development capability.

Familiarity with Development Organization: A SPA is often conducted by personnel from within the organization who are trained in the CMM methodology. As a result, the assessment team is usually very familiar with the organization and may make assumptions about how things work rather than rigorously following the CMM methodology to reach their findings. In contrast, a SCE is conducted by individuals outside the organization who have little, if any, familiarity with processes used within the organization under consideration. Accordingly, the assessment team must be more thorough in their search for objective evidence of process maturity.

The above differences in the conduct of a SPA versus a SCE has led the SEI to say, " By far the most important lesson learned confirmed what we suspected: comparing the results of evaluations to assessments is like comparing apples to oranges, especially when viewed through the maturity-level lens" (Besselman, Byrnes, Lin, Paulk and Puranik, 1993:24).

Another criticism of the CMM is its failure to adequately discriminate between maturity levels. To progress from a lower to a higher maturity rating, *all* the characteristics of the higher level must be met. For example, an organization may exhibit some of the characteristics of a Level 3 process, but the failure to meet all of the requirements of a Level 3 results in a Level 2 rating (Bollinger, 1991:31). As a result, some organizations have informally identified their process maturity in terms of intermediate ratings, such as 1.8 or 2.5. This may reflect the desire by organizations to justify the amount of time and effort invested in the software development process--to show some degree of improvement in their process maturity. Bollinger also states that "...while the SEI process maturity model is clearly intended to help design-intensive organizations become better at developing software, in reality it appears to strongly favor maintenance processes with relatively narrow product definitions" (Bollinger, 1991:27).

In response to perceived limitations of the CMM, and in particular, the SCE as a capability assessment tool, Air Force Materiel Command (AFMC) undertook the development of their own process capability assessment tool. AFMC's perception is that the CMM does not adequately address the systems engineering aspect of software development and is focused on organizational versus program-specific capabilities (ASC, 1993:12). Despite these concerns and the emergence of process maturity model variants, the CMM is still widely regarded as a useful model of organizational software development process maturity.

2.5.4 Other Capability Assessment Methods

The CMM's SCE and SPA are not the only process maturity assessment methods currently in use within the DoD. In 1983, the Aeronautical Systems Center began using the Software Development Capability/Capacity Review (SDCCR) as a tool to "assess an offeror's specific capability and capacity to develop software required on a particular weapons system program as defined in the [Request for Proposal] RFP" (ASC, 1992:1). Unlike the CMM SCE, which is also used for source selection, ASC's SDCCR does not assign a maturity rating, but rather, produces a written report which is incorporated into the final source selection. This report evaluates eight major areas including management approach, management tools, engineering development process, personnel resources, Ada personnel resources, Ada technology, flight critical software, artificial intelligence technology, and complex hardware development (ASC, 1992:9).

Due to perceived shortcomings of both the SCE and the SDCCR, AFMC recently developed a new assessment method, the Software Development Capability Evaluation (SDCE). The SDCE is primarily based on the CMM and the SDCCR and is also used during source selection to determine the strengths, weaknesses, and risks of offerors. It is organized into six functional areas: Program Management, Software Engineering, Systems Engineering, Quality Management and Product Control, Organizational Resources and Program Support, and Program Specific Technologies (AFMC, 1993:3). Unlike either the SCE or the SDCCR, the SDCE recognizes the increasing importance of software engineering in the total systems engineering process.

The SDCE has been approved for use on a few AFMC pilot programs after which the results will be assessed and its applicability for AFMC-wide use will be determined.

2.6 Project Success

What constitutes project success in the software realm? In the general management realm, the parameters of *cost, schedule,* and *quality* figure prominently in the descriptions of project success: A project is usually considered successful when it satisfies project objectives expressed in terms of the three critical parameters of time, cost, and performance; but may include other criteria as well, such as end-item quality (Nicholas, 1990:472). In the software realm, "...the requisites are accurate measures of software cost, schedule, and quality" (Mosemann, 1994:3). In recent years, the parameter of *quality* has taken on particular importance. "Product quality should be the focus of all process improvement" (Hersh,1993:12). "We believe the 1990's will be the quality era in which software quality is quantified and brought to the center of the development process" (Basili and Musa, 1991:91). Thus appropriate measures of software product *success* appear to be the same as those for any other product: *cost, schedule,* and *quality.*

2.6.1 Measures of Cost and Schedule Performance

In order to measure cost and schedule performance, two steps must be followed. First, one must set a performance baseline, and second, one must compare this baseline with actual performance. In project management, the projected rate of funds expenditure (the baseline) is expressed in the Budgeted Cost of Work Scheduled (BCWS). The BCWS can be expressed as the planned expenditure of funds over time, based on the completion of the planned work packages. The Budgeted Cost of Work Performed (BCWP) represents the *earned value* of the work performed, and is an estimate of the work completed (expressed in dollars). The difference between the BCWS and the BCWP is the schedule variance, expressed in dollars, and represents the amount of work which was scheduled, but not performed. The ratio of BCWP to BCWS defines the degree to which a project is ahead of or behind schedule, and is called the Schedule Performance Index (SPI). A SPI of less than 1.00 implies that for every dollar of work scheduled, less than one dollar has been earned--a schedule overrun. A SPI of more than

1.00 implies that for each dollar of work scheduled, more than one dollar of work has been earned--a schedule underrun. A SPI of 1.00 implies an "on target" condition. A third variable used to measure cost performance, is the Actual Cost of Work Performed (ACWP). ACWP is the sum of funds actually expended in the accomplishment of the planned work tasks. Cost Variance is the difference between what the project was expected to cost (BCWP), and what the project actually cost (ACWP). Deviations in the actual versus planned cost can be expressed in the ratio of BCWP to ACWP, and is called the Cost Performance Index (CPI). Similar to SPI, a CPI of less than 1.00 implies that for every dollar of value earned, more than one dollar was actually spent--a cost overrun. A CPI of more than 1.00 implies that for every dollar of value earned, less than one dollar was spent--a cost underrun. A CPI of 1.00 implies an "on target" condition. The indices of CPI and SPI defined above are the standard cost/schedule performance measures for both government and industry (Nicholas, 1990:376-389).

2.6.2 The Effect of Process Maturity on Performance

The value of the performance index CPI indicates whether a project is underbudget, overbudget, or on target. Similarly, the value of the performance index SPI indicates whether a project is ahead of schedule, behind schedule, or on target. Given that the goal of any project is to meet the target budget and schedule, an organization's success can be measured by evaluating the CPI and SPI of a particular project. The closer the CPI and SPI are to a value of 1.00, the more successful the project can be considered, at least in terms of cost and schedule. Thus, it is reasonable to expect that as a organization's process matures, its success or ability to consistently meet target budgets and schedules will increase.

The concept of increasing process maturity resulting in better and more predictable cost and schedule performance can be applied to the CMM's five software process maturity levels (Paulk, Curtis, Chrissis, and Weber, 1993:23). Paulk et al. describe a positive relationship between process maturity and performance. As an organization matures from Level 1 to Level 5, the difference between target results and actual results decreases (i.e., CPI and SPI move closer to 1.00), and the variability of the actual results about the target decreases (i.e., performance becomes more predictable). Graphically, the relationship between maturity and performance can be thought of as a probability distribution whose central tendency at Level 1 is somewhere below the target and whose distribution exhibits a high variance (Figure 2-3). At Level 2, the central tendency of the distribution is now on or very near the target, but the distribution still exhibits a high degree of variance. At Level 3, the central tendency of the distribution is the same as the target, and the variance of the distribution is less than at Level 2. At Levels 4 and 5, the central tendency remains the target and the variance continues to decrease as an organization strives to optimize its process.

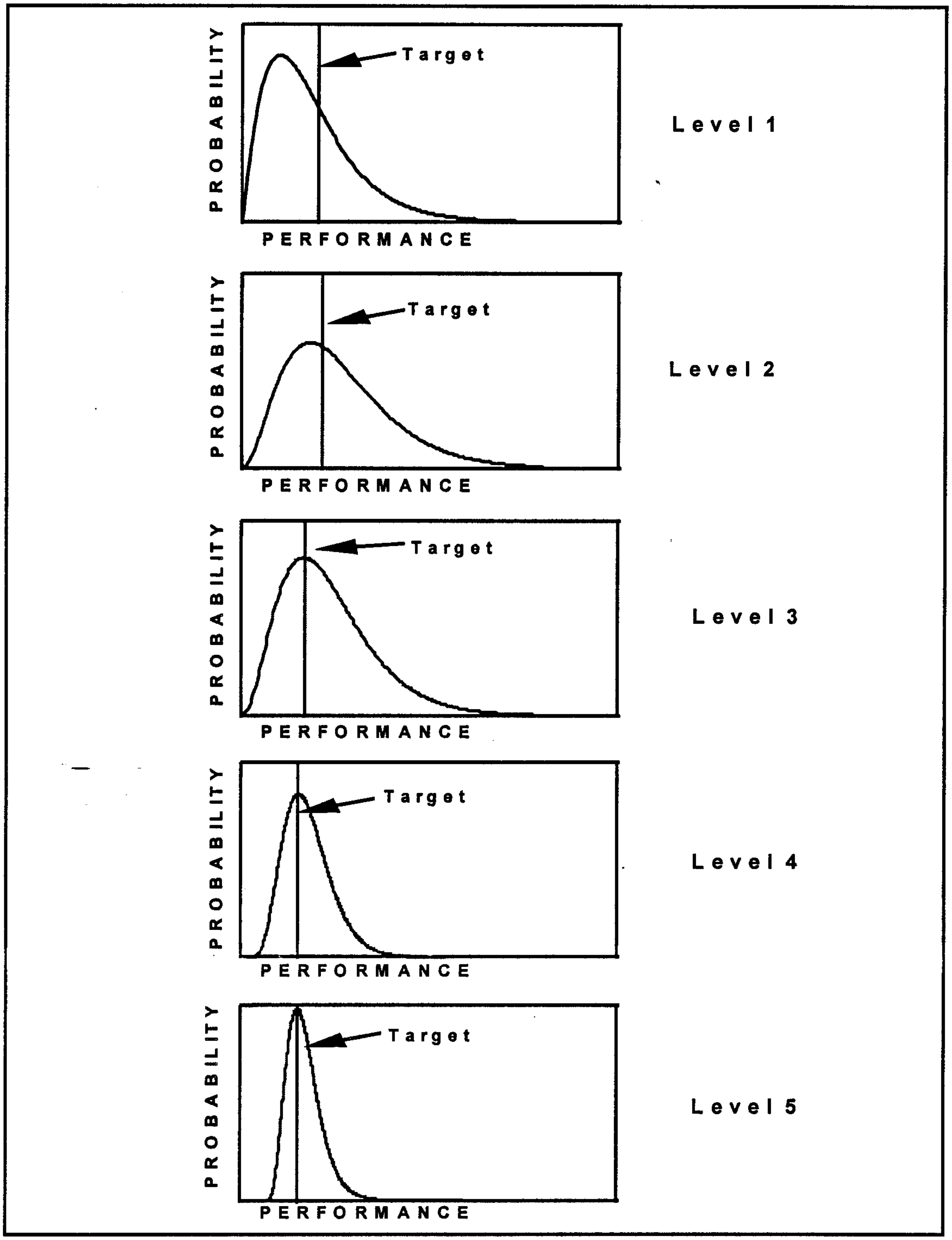

Figure 2- 3 The Effect of Process Maturity on Performance

Adapted from "Capability Maturity Model, Version 1.1," IEEE Software, 10:23 (July 1993).

2.6.3 Measures of Quality

Quality is very difficult to define, much less measure quantitatively. According to Weinberg, "Quality is conforming to some person's requirements...for each person, the same product will generally have different quality...what is adequate quality to one person will be inadequate quality to another" (Weinberg, 1992:5,6). Many measures of quality have been proposed: defect rate, cost, early completion, ease of use, and user satisfaction are but a subset of common quality measures.

"Although defect rates are common measures of quality, quality is not a single idea, but a multidimensional concept" (Basili and Musa, 1991:91). Therefore, no one measure, or limited subset of measures, is globally embraced as the *sine qua non* of product quality measurement. "But whatever the criteria, it is clear that the number of problems and defects associated with a software product varies inversely with perceived quality" (Carleton, Park, and Goethert, 1993:30). Without a universally accepted measure of quality, organizations measure the quality of their product using metrics they perceive as most meaningful. Defect rate, though imperfect as a measure of quality, is relatively easy to measure, is intuitively related to product quality, and thus is not an unreasonable metric for assessing software product quality.

2.7 Anecdotal Evidence

There is accumulating anecdotal evidence supporting the use of the SEI's software process maturity framework as a means to process improvement. These success stories paint an intriguing picture of dramatic improvement and return on investment due to increases in process maturity.

2.7.1 Raytheon

In 1988, Raytheon performed a self-assessment of their Software Systems Laboratory (SSL) division using the SEI's process maturity framework. The SSL rated itself a Level 1. Based on these results, Raytheon initiated a software process improvement program to address the areas of the self assessment that were identified as needing improvement, including policy and procedures, training, tools and methods, and process database (metrics) (Dion, 1993:29).

Between 1988 and 1993, Raytheon invested nearly $1 million per year towards process improvement. This investment moved the SSL from a Level 1 to a Level 3 SEI maturity level. More importantly though, Raytheon estimates that during this period rework costs as a percentage of total development cost has decreased from 41 percent to 11 percent, resulting in an estimated savings of $15.8 million (Dion, 1993:32). In addition, the process improvement initiative has resulted in a two-fold increase in productivity and a $7.70 return on every dollar invested (Dion, 1993:28).

2.7.2 Hughes

In 1987, the Software Engineering Division (SED) of Hughes Aircraft in Fullerton, CA, paid the SEI $45,000 to undergo a software process assessment. The SEI found Hughes SED to be a SEI Level 2, and made the following recommendations:

- establish quantitative process management,
- establish a technical group to be the focus for process improvement,
- review software training requirements,
- insure the SED is involved in the specification development of all new software projects, and
- apply consistent and uniform review practices to the software development process (Humphrey, Snyder, and Willis, 1991:13).

Hughes agreed with these recommendations and implemented an action plan that expended 78 man months of effort over the next two years and cost approximately $400,000.

In 1990, Hughes SED underwent a second assessment which placed the organization at a strong Level 3. As a result of their process improvement efforts, Hughes experienced improved working conditions, higher employee morale, and better cost and schedule performance. The improvement in cost performance was measured by an increase in the CPI from .94 to .97, which translates into an estimated annual savings of about $2 million (Humphrey, Snyder, and Willis, 1991:22).

2.8 Correlational Evidence

Although there exists a growing body of anecdotal evidence that suggests a higher maturity rating results in more successful products, there is currently no established statistical correlation between these two. This lack of statistical correlation is not due to a lack of interest in such correlation: "We're finding that there's not much data out there by which we can measure process-improvement activities" (Hersh,1993:12); "...[CMM ratings] for organizations are so riddled with statistical and methodological problems that it appears unlikely that such ratings have any meaningful correlation to the actual abilities of the organizations to produce high-quality software on time and within budget" (Bollinger and McGowan, 1991:26).

2.9 Summary

The DoD is serious about improving the current state of the software engineering practice. According to Deputy Assistant Secretary of the Air Force, Lloyd K. Mosemann II: "The Pentagon wants:

- Predictable cost.
- Predictable schedule.
- Predictable performance.
- Predictable support and sustainment.

In other words, *predictable quality!*" (Mosemann, 1992:2).

To this end, the SEI's process maturity framework and the CMM have been identified as a means to achieve these goals. This is promulgated in a proposed policy requiring bidders on all Air Force software contracts to have been assessed at an SEI Level 3 or higher by 1998 (Mosemann, 1992:4). This policy will dramatically influence corporate decisions among DoD contractors, and will fundamentally alter the process by which the Air Force contracts for software. Given that this policy is unsupported by empirical evidence, we believe it is valuable to apply a rigorous research methodology to investigate the presumed correlation upon which this policy depends.

3. Methodology

3.1 Overview

Our research involved the collection of secondary, historic data from DoD software development contracts. These data consist of (1) pre-established contractor process maturity ratings (as defined by the SEI's CMM), and (2) cost, schedule, and quality data provided as contract deliverable data. This data was then used to determine if a correlation exists between a contractor's software development process maturity rating and overall software project success. Additionally, moderating data was gathered to enable sample stratification, to gain insight into factors affecting the correlation. Success, for the purposes of our research, is defined by cost, schedule, and quality performance.

Our research methodology consists of four phases: an exploratory phase, a research design phase, a data-gathering phase, and a data analysis phase. These are discussed below.

3.2 Exploratory Phase

The exploratory phase of our research involved review of the relevant literature and discussions with several DoD experts in the field of software development/management. The purpose of this phase was to (1) understand the CMM, how it evolved, and the manner in which it is currently being used; (2) assess the limitations of the CMM, and identify alternative models which are proposed or are currently in use; (3) establish the current state of quantitative analysis supporting the correlation between process maturity and product success; (4) identify the appropriate measures for defining software project success and; (5) determine if quantitative data is available, of such quantity/quality to allow analysis by statistical methods. Items (1) through (4) were addressed in the previous chapter. Item (5) is addressed in Chapter 4.

3.3 Research Design Phase

The goal of the research design phase was to establish a research design which answers the research question "does a correlation exist between process maturity rating and software project success?" The CMM has gained acceptance because of the intuitive and anecdotally-supported understanding that a more mature software development process will, as a matter of course, produce better software. To validate this presumed correlation, it would be meaningful to analyze the historical record and determine if the presumed correlation is statistically confirmed.

In order to establish a correlation having statistical validity, a large body of data must be gathered. This amount of data exists in an historical context. There is a wealth of secondary, historical data, generated as a by-product of the DoD software procurement process. This data is in a relatively consistent format (often mandated by government standards), and was available to us via the procuring organization.

By gathering historical contract data, we have taken advantage of several notable characteristics which may not apply to other forms of secondary data. These characteristics are validity, relevance, and reliability. The validity of contract data is defined by the degree to which it adequately describes contractor performance. Project managers define project success in terms of cost, schedule, and quality, as reported by contract data. Thus cost, schedule, and quality data are relevant and valid from the project manager's perspective. The reliability of contract data is enhanced because the collection, content, and reporting are governed by DoD guidelines, and because the same criteria for cost and schedule measurement and reporting are mandated across all contracts. These guidelines and criteria, known formally as Cost/Schedule Control Systems Criteria (C/SCSC), establish a generally consistent format across all government procurements, thus allowing comparison of data from different contractors and different contracts.

Contract data is not perfectly homogeneous and perfectly consistent, however. As the designation indicates, C/SCSC is a set of criteria for cost/schedule progress measurement and reporting. It does not impose a "standard" cost and schedule control system, but rather, defines a set of minimum standards for the cost and schedule management systems used by government contractors (Christensen, 1993:7). Thus, variations can be expected between contractors' reporting systems which may not be fully accounted for in our methodology. Some of these variations will manifest themselves in the cost and schedule performance indices, where others may not. Part of our analysis includes evaluation of outliers, and analysis of clusters to determine if some confounding effect, not captured by our initial methodology, is at work.

3.3.1 Selection Criteria for Sample

An important criterion for sample selection was that the contractor had been rated using the SEI's CMM. The CMM has been in existence since 1987. During this time, is has been applied to DoD software contractors on a limited basis. Therefore, the population of contractors which have been rated is relatively small. The available data set was further restricted by the following criteria:

1. Contractors must have produced software for the DoD within the same timeframe as the SEI rating (for the purposes of this study, from six months prior to six months after the rating is established).
2. Above procurements must have reported cost/schedule data per C/SCSC.
3. Costs must have been reported in sufficient detail as to identify software specific efforts.

We recognized software quality data, unlike cost and schedule data, is not collected in a consistent format, specified by government-imposed criteria. Therefore, we did not reject potential respondents based upon the absence of consistent quality data.

The Cost Libraries at ASC and Electronic Systems Center (ESC) provided an excellent means to rapidly identify potential data points. At the Cost Libraries, we could quickly identify those programs which met constraints two and three above. Since cost performance reports (CPRs) are required for contracts which comply with C/SCSC, we were able to quickly identify programs which met criterion two (C/SCSC reporting) by searching through the library of CPRs. By identifying software-specific work breakdown structure (WBS) elements during this search, we also identified those programs which met criterion three (software specificity). Thus, the Cost Libraries enabled us to restrict our direct inquiries of the program offices solely to the matter of SEI rating and criterion one (temporal association).

During our consultation with experts, we were exposed to general pessimism that we could obtain a data set of sufficient size upon which to perform valid analysis. According to the consensus, relatively few contractors were rated, and of those, few could be expected to report software development costs to sufficient detail as to be distinguishable from non-software efforts. As we report in Chapter 4, our net data yield did not conform to the consensus. We found sufficient data for this study.

3.3.2 Quantitative Measures

An objective of our research was to gather indicators of software project success. As previously established, cost, schedule, and quality are generally accepted as measures of success. Accepted Government and industry-wide standard measures of cost and schedule performance are the cost performance index (CPI) and the schedule performance index (SPI). These indices are reported in the CPR, and are defined as follows (see section 2.3.1 for further detail on how these measures are derived):

CPI = BCWP/ACWP

SPI = BCWP/BCWS

where: ACWP = Actual Cost of Work Performed.

BCWP = Budgeted Cost of Work Performed.

BCWS = Budgeted Cost of Work Scheduled .

Quality measures are not as universally-accepted as those for cost and schedule. One measure, software defect rate, is typically gathered to monitor software development progress during the coding and test phases. Defect rate is defined in terms of the number of software defects or errors per quantity of code generated, typically expressed in errors per thousands of source lines of code (KSLOC). Different organizations may define the terms "error", and "source lines of code" differently. Furthermore, this data may or may not be formally reported to the government, and the format may vary. In some cases, the data was reported and tracked at the Defense Plant Representative Office, and only extraordinary variances from the established norms or targets were reported to the program office. For our research, we gathered quality data in varying formats. However, for many of the sample programs, data on product quality was not available, or was in such a format that was difficult to legitimately normalize. Thus, we drew our project success conclusions solely upon the basis of cost/schedule information.

3.4 Data Gathering Phase

The following are the general steps we used to collect data. We started with a representative sample of contractor organizations, collecting data on relevant projects, both from the appropriate cost library, and from the program office. The steps can be summarized as follows:

- Identify contracts which report software development costs as a discrete contract work breakdown structure (CWBS) element.
- For each contract identified, establish whether the contractor has been rated per the CMM methodology.
- Collect cost/schedule information for timeframes relevant to the ratings.
- Collect moderating data which may be used to characterize the software development project--to enable sample stratification.

The steps are depicted in figure 3-1.

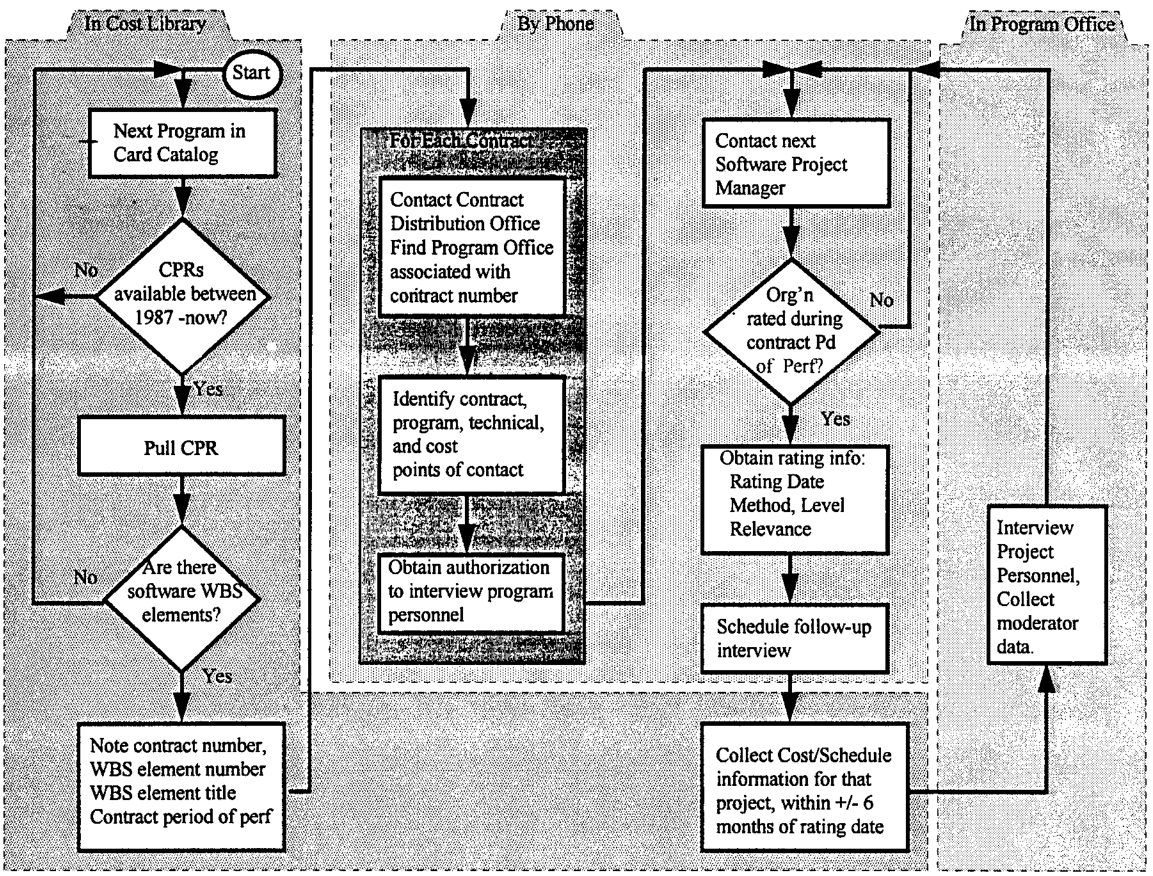

Figure 3- 1 Data Gathering Flow Chart

The steps in figure 3-1 are elaborated in the subsequent sections. It is significant to note that the linear nature of the data collection activity is representative of the logical flow of the data identification and collection process. Efficient data identification and collection required that, throughout the process, multiple candidate programs would be in various stages of the data identification and collection pipeline at any given time.

3.4.1 Software Development Project Identification

Programs which track cost and schedule progress per the C/SCSC criteria are required to archive historical cost performance reports (CPRs) at the Cost Library for their product division. This represents a rich resource for cost/schedule information, which can be efficiently scanned for programs which meet the cost reporting criteria established above. By reviewing archive data we were able to identify contracts which report software development costs as a discrete contract work breakdown structure (CWBS) element.

At the ASC library, we searched the catalog of current programs alphabetically, pulling the CPRs for each program which reported within the 1987-present timeframe. Prior to 1987, the programs would not have been rated by the CMM methodology. Examining the CPRs, we quickly determined, by the titles of the contract work breakdown structure, if software costs were reported as distinct elements. If a CWBS dictionary was available, we checked the dictionary definition to verify the element was exclusively software. We noted those programs with software-distinct CWBS elements, logging the program name, the contract number, and the numbers and titles of the CWBS elements that appeared to be software-related.

Our methodology was somewhat different at the ESC library, since the data had to be gathered during the course of a two-day temporary duty visit. We obtained a list of candidate programs from a point of contact (to remain anonymous) at ESC. With this list, we searched the database at the ESC Cost Library to identify which of these programs reported software costs as distinct WBS elements. From that point, the identification of candidates was the same as that employed at ASC.

3.4.2 Contractor Rating Verification

For each contract which reported software development costs as discrete CWBS elements, we contacted the responsible systems program office (SPO). The purpose of contacting program personnel is twofold: first, permission to gather data for research purposes must be obtained. Second, the program personnel must provide valid SEI rating information, including the date of the rating, the method by which the rating was obtained, and other relevant program information which would provide additional insight into the program (moderating variables).

We traced the system program office through the contract number identified on each CPR, and eventually got in contact with the appropriate program personnel. The process of finding the appropriate knowledgeable personnel, and getting into contact with them to obtain rating information and moderating information was the most challenging aspect of the data gathering phase. The first problem was finding the right person. Since our data involved technical, programmatic, and cost information, we had to find the technical project officer for software development, as well as the program control person. To reach these people, we often had to convince the program director to allow us to talk about program issues with their personnel. Many programs are very sensitive to disclosing their cost and schedule performance to alien agencies. In many cases, we had to dispel concerns about disclosing potentially inflammatory information about the program, and went to great lengths to assure their program anonymity.

Some project managers did not know of the CMM, or their contractor's ratings, while some were very familiar. In most cases, we obtained rating information directly from the project manager, although in some cases the project manager had to contact the contractor to get CMM rating information. In several cases, we had to contact the government's SETA (Scientific, Engineering, and Technical Assistance) contractor for project information. In one case, with the project manager's authorization, we contacted a rated contractor directly. It is important to note that we did not attempt to independently verify the rating information provided by the program personnel.

3.4.3 *Rating-to-Project Relevance*

For each project, the validity of the correlation between the CMM rating and project cost/schedule performance depends upon whether the project under consideration was used in the CMM rating process. This associative relevance was deemed a necessary moderator to account for the degree of association the rating had with the project under consideration. Four scenarios define the four degrees of rating-to-project relevance:

1. Very High Rating-to-Project Relevance--the project under consideration was itself rated using the CMM rating process. Thus the organization's rating is based solely upon the project under consideration.
2. High Rating-to-Project Relevance--the project under consideration was one project of several used in obtaining the CMM rating for the organization.
3. Medium Rating-to-Project Relevance--the project under consideration was not used to establish the CMM rating, but the organization or personnel which participated in the project were also responsible for projects evaluated in the CMM rating of interest.
4. Low Rating-to-Project Relevance--neither the project, nor the personnel responsible for the project under consideration were used to obtain the organization's CMM rating. In this case, the rating for the contractor as a

whole is considered to apply to the organization responsible for the project under consideration.

We recognized that programs with medium and low rating-to-project relevance may adversely affect the validity of the correlation between rating and performance. At the outset, our concern for the the scarcity of data militated against eliminating medium and low relevance projects from consideration. Instead, characterizing the relevance of the data enabled sample stratification which enabled us to account for any relevance-related effects.

3.4.4 Collection of Success Indicators

For each contractor identified as having been rated per the SEI CMM methodology, whose program tracks software development costs by discrete CWBS elements, and whose rating date(s) has been identified, we were then able to gather CPI and SPI data relevant to the rating timeframe. As previously stated, CPI and SPI are derived from cost and schedule data, available in CPRs maintained in the ASC and ESC Cost Libraries. To get data that is representative of the contractor's performance which contributed to the CMM rating, we gathered cumulative cost/schedule data from three and six months prior to and after the rating date. By taking the difference between the cumulative cost/schedule data collected six months after the rating and six months prior to the rating, we were able to calculate performance indices that are representative of a twelve month period of performance. This temporal linkage should provide good correlation between the CMM evaluation and contractor cost/schedule performance, while also providing, to some degree, attenuation of any month-to-month variability in the performance indices. Note that cost/schedule data was collected at three and six month intervals prior to and after the rating in order to provide insight into the behavior of the cost/schedule data over the period of interest. However, performance index calculations were derived from the data taken at six months prior to and after the rating date, giving a 12-month "snapshot" of cost/schedule performance.

Quality metrics, the third success indicator, is not reported in the CPRs, and had to be collected during the project manager interview. Until recently, there has been no formal policy directing the standardized measurement and reporting of quality metrics (Mosemann, 1994:3). We acknowledged that there would likely be a variety of quality metrics reported, and we attempted to obtain some common criteria, such as Defects per Standard Line of Code or Defects per Module. However, there was no consistency to the quality data, and normalization was not possible. Thus the success determination was derived from cost and schedule data alone.

3.4.5 Moderating Data

Although cost and schedule data are intended to characterize projects success, other moderating factors may affect cost and schedule performance. In the course of our literature review and exploratory phase, we identified several factors which may influence project success.

We conducted personal interviews with program and technical managers of each project using a standardized data collection form to guide the interviews and record data pertaining to the moderating variables. During the course of the interviews, some moderators had to be deleted because, in our view, they provided too close a link to the project, and thus would compromise the anonymity of the respondents. Acquisition category was such a moderator. We felt that a reasonably knowledgeable reader could deduce the identity of a respondent by comparing the acquisition category, application type and year of rating. We felt that these three data in combination could uniquely identify a respondent. The final set of moderators used in the analysis is given in table 3-1.

Table 3- 1

Moderators

Moderating Variable	Description	Rationale
Rating Type	SPA (internal) SPA (external), SCE	Method of rating determination is believed to affect rating level.
Rating Relevance	Low, Med, High, Very High	Relevance of rating could affect reliability of rating
Acquisition Phase	RDT&E, Production, O&M	Different phases of acquisition may affect cost/schedule.
Contract Type	Fixed Price; Cost Plus; Incentive/Award Fee	Different contract/fee arrangements may affect cost/schedule.
Software Lifecycle	Requirements; Design; Code; Test; Support	Different phases of software lifecycle may affect cost/schedule.
Language	Ada, Fortran, Jovial, C++, Other	Programming Language used may affect cost/schedule
Language Percentage	Percentage of project coded in dominant language	Programming Language used may affect cost/schedule
Application Type	Avionics, Command & Control, Database, Simulation, Other	Different application types may affect cost/schedule
Project Budget	Budget at Completion	Monetary size of project may affect cost/schedule.
Budget Volatility	Low, Medium, High	Uncertainty/reduction in funding may affect cost/schedule.
Size	Lines of Code (LOC)	Size of program may affect cost/schedule.
Percentage New Code	Percentage New/Modified Code	Percentage of new/modified versus reused/lifted software may affect cost/schedule
Requirements Volatility	Low, Medium, High	Uncertainty/changes in project requirements may affect cost/schedule
Rebaselining	Rebaseline during period of interest? Yes/No	Changes in program baseline may affect cost/schedule data
Quality Standards	On contract? Yes/No	Quality standards on contract may influence procurement
Quality Parameters	Reported to Program Office? Yes/No	Quality parameters reported to the program office may influence procurement
Program Activity	> .01 of budget expended over 12 month period	Programs with little activity may skew CPI and SPI numbers
Percent Complete	< 25% BAC expended 25% to 75% BAC expended > 75% BAC expended	Stability of CPI and level of SPI are affected by the percentage complete
Baseline Volatility	< 80% Change in BAC > 80% Change in BAC	Stability of the program baseline may affect cost/schedule data

3.5 Data Analysis Phase

CMM rating data is at best ordinal in nature. Hence; statistical analysis techniques such as multiple linear regression, which require interval or ratio data, cannot be rigorously applied. However, a combination of descriptive and non-parametric techniques are adequate to establish the presence or absence of a statistically significant correlation of software development process maturity and software product success. Moderating variables were used to stratify the sample to obtain insight into the factors affecting the correlation of the CMM ratings with cost and schedule data. The results of the analysis is presented in Chapter 5.

Some of the tools to be used in the course of the analysis include:

- Scatter Plot of the dataset: CPI and SPI versus Rating
- Histogram of the frequency density for each rating level
- Box and Whiskers plot of the dataset
- Wilk-Shapiro evaluation of normality at each level
- Kruskal-Wallis Test
- Multiple Comparison Test
- Descriptive Statistics

The first three techniques help visualize the relationship between the rating and performance indices while the latter four provide quantitative results allowing objective comparisons. The graphical techniques for nonparametric analysis are common and relatively intuitive. The Kruskal-Wallis nonparametric analysis of variance and the multiple comparison test are less familiar, and are explained below.

The Kruskal-Wallis Test is a nonparametric analysis of variance that tests the null hypothesis that samples subjected to different treatments (i.e. CMM ratings) actually belong to the same population and therefore would have the same median performance index. The alternate hypothesis would suggest that performance indices at the different CMM rating levels are in fact distinct populations. The rejection of the null hypothesis thus would suggest that there is a difference in the median performance of organizations at different maturity levels.

In order to test the null hypothesis, the sum of the ranks R_j for each sample must be obtained. This is done by ranking N, the total number of observations from 1 (the smallest performance index) to N (the largest performance index) and summing the ranks within a sample. When the null hypothesis is true, all observations come from the same population and we expect the ranks to be equally likely distributed between the samples. If, however, the null hypothesis is false, then some samples will consist mostly of observations having small ranks (lower performance indices), while others will consist mostly of observations having large ranks (higher performance indices). The sum of the ranks R_j for each sample can then be used to calculate the Kruskal-Wallis H test statistic according to "Equation (1)" below.

Kruskal-Wallis H Test Statistic

$$H := \frac{12}{N \cdot (N+1)} \cdot \sum_{j=1}^{k} \frac{\left[R_j - \frac{n_j \cdot (N+1)}{2}\right]^2}{n_j} \qquad (1)$$

where

k = number of samples

R_j = sum of the ranks in the jth sample

n_j = number of observations in the jth sample

N = total number of observations

By referring the value of H to the chi-square distribution, a P-value can be found and used to accept or reject the null hypothesis. The P-value is the probability that the distributions appear to be distinct when, in fact, they are not. For our analysis, we used the statistical analysis software package, Statistix 4.0, to calculate the Kruskal-Wallis test statistic. This test was performed at a significance level of .05, meaning conclusions can be drawn with a 95% level of confidence.

The Kruskal-Wallis Test can only determine if at least two of the samples are from different distributions. In order to determine if there is a statistically significant difference in more than one pair of samples and which samples differ from which others, a multiple comparison test is required. Using the multiple comparison inequality, "Equation (2)" below, Multiple Comparison Inequality

$$\left|\overline{R_i} - \overline{R_j}\right| \le z \cdot \sqrt{\frac{N \cdot (N+1)}{12} \cdot \left(\frac{1}{n_i} + \frac{1}{n_j}\right)} \qquad (2)$$

where

$\overline{R_i}$, $\overline{R_j}$ = mean rank of the ith and jth sample

n_i, n_j = number of observations in the ith and jth sample

N = total number of observations

z = 1.834 at a level of significance of .20

we compare the difference in mean ranks for two samples, where the mean rank, $\overline{R_j}$ is simply the sum R_j of the ranks in that sample divided by n_j, the number of observations in that sample; that is, $\overline{R_j} = R_j / n_j$. If the absolute value of the difference of the mean ranks between two samples is less than the right-hand side of "Equation (2)," then the null hypothesis is true and there is no significant difference in the samples under consideration. However, if the absolute value of the difference of the mean ranks is greater than the right-hand side of "Equation (2)," then the null hypothesis is false and there is a significant difference in the two samples.

The multiple comparison test was performed at a level of significance of 0.2 which implies a 80% level of confidence in the result. It is important to note that the overall level of significance used in multiple comparisons are frequently larger than those ordinarily used in an inference involving a single comparison. The level of significance chosen for our analysis is consistent with the values recommended (0.15 to 0.25) for this type of nonparametric analysis technique (Gibbons,1976:182).

3.6 Methodological Difficulties

Our research methodology is based upon the collection and analysis of historical data. This methodological approach requires the availability and consistency of data. Our focus on cost and schedule data provided by standard means helped reduce the potential error in our primary data set. However, the same standards could not be applied to the moderating data obtained to characterize the software projects from which the cost/schedule data were derived. The lack of standards for these moderating variables may have affected both the consistency and the validity of these moderators.

A degree of subjectivity, and researcher bias was unavoidable. Researcher bias was introduced primarily by our selection of moderators. Since it was impossible to fully characterize the software development environment for each respondent, we were forced to select a relatively small set of moderators we felt would provide the most meaningful information. We selected these moderators based on our literature search, discussions with experts, and educational experience. Lacking the resources to perform in-depth case studies of all respondents, we were compelled to rely upon the insight our moderators provided to characterize the context in which the cost, schedule, and rating data were derived. This bias effect introduced by moderators was combined with the subjectivity of the respondents in providing the values attributed to the moderator variables. In most cases, our respondents relied on their best judgement when responding to our moderator-related inquiries. Although the interviewees were qualified to provide this insight, there is no guarantee that they provided accurate or complete information.

A significant compromise to our initial methodology arose when project quality data failed on both the consistency and availability criteria. We found that the data to enable an analysis of the quality characteristics of the software projects was sparse, and of widely varying format. We felt the data would have provided no meaningful comparisons at any quantitative level, and was thus simply reduced to bi-level moderators.

3.7 Consistency of Moderating Variables.

Some of the moderating variables we intended to collect proved difficult to reliably obtain. For example, respondents in some cases failed to distinguish between "new" and "modified" code. Thus we had to modify our moderator to accommodate this lack of distinction. The moderator "Percent New/Modified Code" enabled us to distinguish between code which required significant design and engineering, and code which was reused or "lifted".

With regard to language distinction, some respondents didn't distinguish between the amount of code written in a variety of higher order languages. For example, if a project consisted of the languages Ada, Fortran, and Assembly, the program may have only reported the amounts of code in terms of "HOL" (higher-order language) and Assembly. Thus, in some cases we were unable to identify the code as being predominantly one language.

The moderator project size should only be used to distinguish projects whose size differs by an order of magnitude. Different definitions of lines of code (DSI vs. KSLOC) combined with different languages/combinations of languages seriously degraded the absolute accuracy of the program size data collected. In using this moderator we chose to stratify on the arbitrary, but commonly recognized breakpoint of 100,000 lines of code, separating the data set into roughly balanced subsets.

The distinction between application types was subjective. In many cases, a particular project encompassed various application types, and it was left to the program personnel interviewed to characterize the project into a type which best fit the project. Due to the subjectivity of the application types, we chose to stratify on the gross distinction between "real-time" and "information system" applications. This was to capture the relative complexity of these broad categories.

Some moderators could not be gathered with consistency, due to an inherently subjective nature. For example, the moderator "requirements volatility", used to characterize the degree to which the requirements changed during the course of the project, was strictly based on the expert opinion of the program personnel interviewed. Lacking an objective baseline or comparator, the interviewees' perceptions may have varied widely from one program to another. Thus we were unable to derive significant conclusions from this moderator.

3.8 Summary

The objective of this thesis was to determine if a correlation between SEI CMM rating and software product success exists. This objective was met in a four-phase manner: (1) Exploratory Phase, (2) Research Design, (3) Data Gathering Phase, and (4) Analysis Phase.

4. Results

4.1 Overview

This chapter discusses the results of the data gathering, explains the process of identifying and eliminating erroneous data points from the database, and describes the nature of the resulting database. In the process, we discuss some key concepts and definitions pertaining to the data, the process of reducing and conforming the data, the derivation and coding of performance indices and derived moderators, the nature of the dataset, and finally, a description of the final database to be analyzed.

4.2 Concepts and Definitions

For the purposes of this discussion, a "data point" is defined as an instance, or set of circumstances where, for a given software development project, rating data and cost data exist, and are mutually relevant. This is the set of circumstances whereby

1. the software development project reports cost and schedule data per the guidelines of C/SCSC
2. the organization conducting the software development project has been rated in accordance with the SEI CMM guidelines
3. The cost and schedule data are representative of the rating timeframe, and

4. The rating is generally representative of the project for which cost and schedule data are reported.

Based on the above definition, it is clear that multiple data points may arise from a particular organization, program, or project. In general, an individual organization may have multiple programs which fall within our sampling criteria. Additionally, each program may have one or more projects which meet the sampling criteria, meaning that the cost and schedule data were reported for software-unique work packages or projects. Finally, each individual project may have been in progress during multiple rating periods, and thus would provide cost and schedule data relevant to each rating period.

An example of this is shown in figure 4-1, for an organization (DoD contractor) which has been rated twice and has two programs (Government contracts A and B), one of which has three individual software projects (WBS elements), the other only one. Note that these Government contracts have different periods of performance: Projects one through three of Program A were in effect for two rating periods, whereas Project one of Program B was in effect for the last rating period only. In this scenario, this one organization would have provided seven discrete data points.

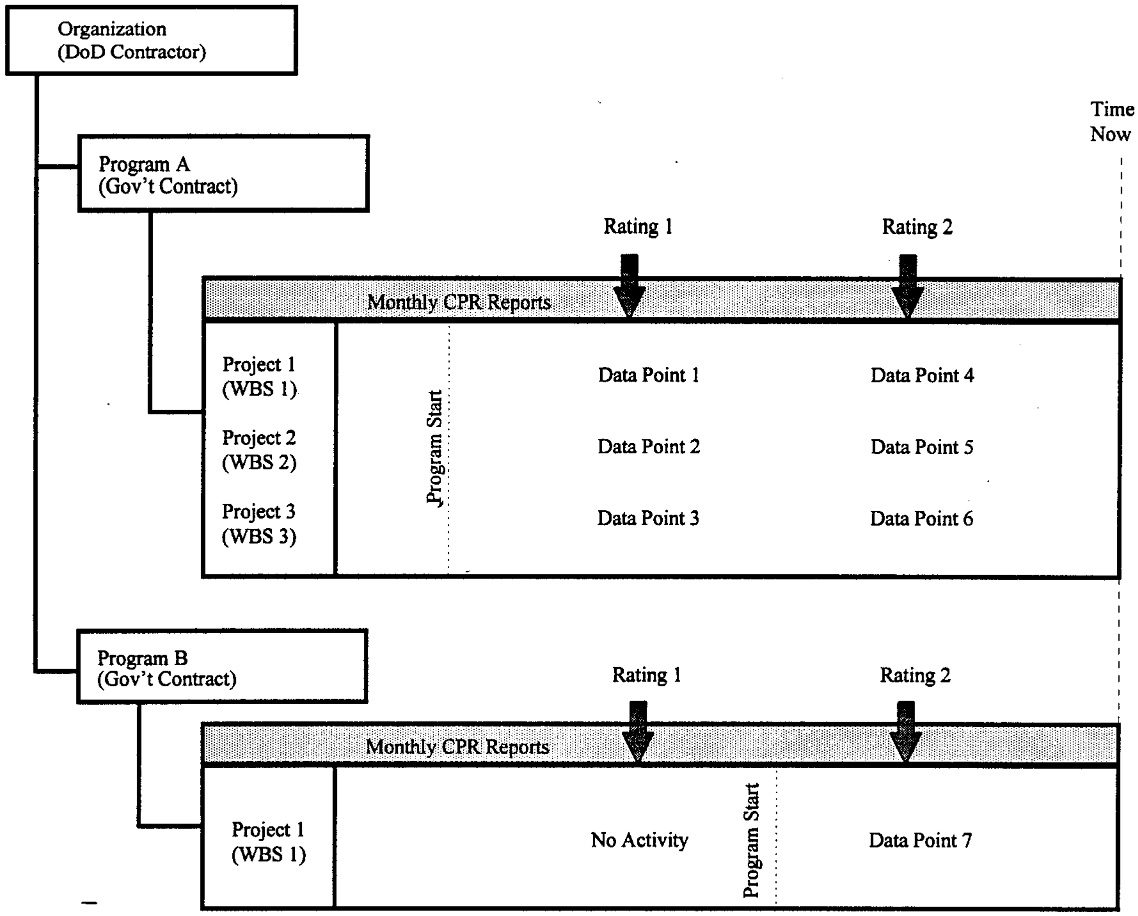

Figure 4- 1 Origin of Data Points

Each data point can be represented by two ordered pairs of rating and performance index, and plotted on a coordinate system. Note that we calculate both the SPI and CPI, so each data point will be characterized by both indices.

Each data point is also characterized by other parameters which lend context to the data point. These parameters are called moderating variables, and may provide insight into the factors that influence the correlation between the performance indices and the ratings.

4.3 Data Reduction

Based on the requirements and constraints set out by the research design and methodology, as well as the opinions of the experts we interviewed during the exploratory and design phases, we expected a small sample of data points. The unexpectedly large number of data points made the automation of the data analysis a requirement. Therefore, considerable effort was invested to reduce the data to a database-compatible format, so that efficient analyses could be performed. The data were collected using a standardized data collection form shown at Appendix A. The data collection forms underwent some modifications during the course of the inquiry, as a result of our evolving understanding of which moderators were actually significant, and which moderators could be reliably obtained from the personnel interviewed. The final set of moderators was provided in Table 3-1 of Chapter 3.

The data collection form was designed so that program identification information could be disassociated from the rest of the data to ensure anonymity of the data source. After the data collection forms were completed, the program identification information was separated and secured. Only the researchers and their faculty advisors have access to the correlation matrix which links these programs to their data points. The format of the data collection form was determined prior to the decision to automate the data analysis. Thus the correlation between the data collection form and the database is not exact.

The completed data collection forms were transcribed into a database (Microsoft Access version 2.0). The database was constructed in a flat file format, with each database record representing an individual data point, comprised of identifying code, rating information, cost/schedule information, and moderating characteristics. Each record (data point) in the database consists of fifty-one fields, broken up as follows:

1. Three fields of primary key identifiers. Each data point is uniquely identified by a three-character alphanumeric designator, which identifies the program, rating in sequence, and the WBS element in sequence. This coding scheme allows unique identification of the data point without divulging the identity of the contractor or the contract.

2. One comment field for WBS description. This description is generic, to describe the sort of task the WBS represents, but not to identify the program.

3. Five fields of rating information pertaining to every WBS in a given rating period, including a comment field for comments relating to the rating.

4. Three fields for moderating data related to the program (of which the individual WBS is a part). Moderator fields span rating and WBS domains, and include a comment field for comments relating to the program.

5. Fourteen fields of project-related moderating information. These data relate to the specific WBS (project) being evaluated, and include a comment field for program manger comments relevant to the analysis.

6. Twenty-four fields of WBS-element-specific cost and schedule information. These include the cost/schedule parameters, BCWS, BCWP, and ACWP. Also

recorded are the dates of the data, the budget at completion (BAC) and the latest revised estimate (LRE) for the WBS. These data are used to calculate the SPI and CPI, as well as other moderating data, such as percentage of project completion, and degree of project activity.

7. One field for investigator comments. This provided us with a way to characterize the data point in terms of its relevance to the analysis.

The above descriptions characterize the database. The contents of the database is provided at Appendix B. Appendix B also contains data on derived moderators and performance indices derived from the database.

4.3.1 Conforming the Database.

Although 63 data points were originally collected, only 52 were used in the analysis. We excluded 11 of the data points on the basis of a lack of contract effort during the period of interest. The reason for this exclusion is that the non-cumulative performance indices we measured become extremely unstable at low levels of contract effort (recall that performance was measured over the 12 month period surrounding the rating date, and therefore were non-cumulative). The instability is due to the fact that if little effort is expended on the contract, actual costs (ACWP) during the time period are small, causing CPI to be extremely sensitive to relatively minor variations in earned value taken (BCWP). Likewise, if little work over the period is planned per the baseline (BCWS), small fluctuations in earned value taken (BCWP) can result in large fluctuations in SPI.

Thus, at low levels of contract activity, SPI and CPI become more sensitive to random "noise" in the accounting system than to real variations in contractor performance.

Screening for contract activity on a given project was accomplished by calculating a ratio of contract activity during the twelve month period (twelve-month change in the parameter) relative to total activity to date (cumulative value of the parameter). This ratio was calculated for the three parameters, BCWS, BCWP, and ACWP. If any one of these parameters exhibited an activity level of less than 1%, it was excluded from the dataset.

The resulting dataset is referred to in the research as the "Complete Dataset." The Complete Dataset is to be distinguished from the "Gross Dataset" which encompasses all data taken, regardless of project activity over the period of interest. The Gross Dataset is provided at Appendix B. Those datapoints excluded for whatever reason are flagged with an appropriate comment in the "Investigator Comment" field of the data form. The Complete Dataset is the set from which all subsequent analysis within this study was performed. The comparison of both datasets is presented in Table 4-1.

Table 4- 1

Comparison of the Gross and Complete Datasets

Criteria	Gross Dataset Count	Complete Dataset Count
Number of Contractors	11	11
Number of Programs (Contracts)	14	13
Number of Projects (WBS Elements)	33	31
Number of Data Points	63	52
Number of Data Points from ESC	45	35
Number of Data Points from ASC	18	17
Average Number of Data Points per Program at ESC	9	7
Average Number of Data Points per Program at ASC	2.0	2.1
Average Number of Data Points per Program	4.8	4.0
Average Number of Data Points per Project	1.9	1.7
Average Number of Ratings per Contractor	1.9	1.9

4.3.2 The Nature of the Complete Dataset

In order to obtain the clearest picture of the nature of the relationship between the performance indices and rating, the dataset had to be large enough to be statistically significant, and representative of the relevant population.

The size of the dataset is critically important for any statistical analysis to be valid. If we had been able to collect only five or six data points at one or two rating levels, the validity of our correlational analysis would be highly suspect. Fortunately, the mass of historical data was sufficiently large, and we were able to net 52 individual data points over three rating levels (17 at Level 1, 18 at Level 2, and 17 at Level 3). To improve the likelihood that the data were representative of the relevant population, we collected data from two product centers, the ASC, at Wright-Patterson AFB, OH, and the ESC, at Hanscom AFB, MA. Since we are most interested in software-intensive programs, these two product centers are reasonable candidates to provide samples of our relevant population. Since we expected only a few programs to fit within our sampling criteria, we did not conduct extensive analysis to ensure a representative population. Our goal at the outset was to collect everything which met our sampling criteria, and evaluate the nature of the sample after collection.

We obtained approximately twice as many data points from ESC as from ASC, even though we evaluated fewer programs at ESC. We found that on average, ASC had fewer software specific projects per program than did ESC, not surprising given the nature of the work performed at ASC and ESC. At ASC, software is typically a part of a subsystem on an aircraft-related program, whereas at ESC, software comprises proportionally more of their electronics-related programs.

Table 4-2 expresses some of the characteristics of the complete dataset. Although this sample is probably not truly representative of programs throughout the DoD, this dataset is presented, for the purposes of this research, as a generally representative sample of the population of interest.

Table 4- 2

Characteristics of the Complete Dataset--Count by Moderator

Characteristic	Count
Number of Programs	52
Number of Projects Rated Level 1	17
Number of Projects Rated Level 2	18
Number of Projects Rated Level 3	17
Number of Projects with High-Very High Rating Relevance	40
Number of Projects with Med-Low Rating Relevance	12
Number of Projects Rated using a SPA	34
Number of Projects Rated using a SCE	18
Number of Projects with Less than 15% Baseline Volatility	38
Number of Projects with Greater than 15% Baseline Volatility	14
Number of Projects with Cost-type Contracts	17
Number of Projects with Fixed Price-type Contracts	21
Number of Projects less than 80% Complete	21
Number of Projects greater than 80% Complete	31
Number of Projects Implementing Real-time Applications	25
Number of Projects Implementing Information System Applications	26
Number of Projects Implemented in Ada	24
Number of Projects Implemented in Non-Ada	19
Number of Projects less than 100K LOC	21
Number of Projects greater than 100K LOC	17

4.4 Derivation of Performance Indices and Derived Moderators

Cost performance Index (CPI) and Schedule Performance Index (SPI) are derived from the cost/schedule data obtained from the cost performance reports. Derived moderators are moderators which result from combinations of the parameters ACWP, BCWP, BCWS, BAC, and LRE, already present in the set of cost/schedule data. These derived moderators include:

1. Baseline Volatility--the ratio of the change in the BAC during the twelve-month period to the BAC at the beginning of the period
2. BCWS Activity--the ratio of the change in BCWS during the twelve month period to the total BCWS at the end of the period
3. BCWP Activity--the ratio of the change in BCWP during the twelve month period to the total BCWP at the end of the period
4. ACWP Activity--the ratio of the change in ACWP during the twelve month period to the total ACWP at the end of the period
5. Percent Complete--the ratio of BCWP at the end of the period to the BAC at the end of the period

Performance indices, as well as derived moderators, were not incorporated into the database itself for reasons of limiting the database size. Instead they are calculated by means of queries executed on the dataset. The output of a query on a dataset is another dataset which contains the results of the query operations. In our case, our queries calculated the performance indices and other derived moderators. It is this output which we analyzed, the results of which are found in chapter five--Analysis.

4.5 Coding of Moderators

Moderating variables which are not categorical in nature had to be coded in order to be efficiently analyzed. For example, the moderator "Size" had to be resolved into the levels "Small," and "Large" based on some coding scheme. We analyzed only a small subset of the moderators we collected and coded. We coded all moderators regardless of whether they were incorporated into the current analysis, to facilitate future analysis of this dataset. The stratification and coding schemes are outlined in Table 4-3.

Table 4- 3
Moderating Variables and their Stratification Schemes

Moderating Variable	Range, Levels	Stratification/Coding Scheme
Rating Type	SPA (Int), SPA (Ext) SCE	SPA: Software process assessment. SCE: Software Capability Evaluation.
Rating Relevance	Low, Med, High, Very High	High/Very High: Projects were used to obtain the organization rating.
Acquisition Phase	Concept Exploration R&D, EMD Production, Support Post Release Support	Pre-Production: Concept Exploration, R&D, EMD, Post-Production: Production, Support, Post Release Support.
Contract Type	Firm Fixed Price, Fixed Price Incentive Firm Tgt, Fixed Price Award Fee, Cost Plus; Incentive/Award Fee	Fixed Price: Includes FFP, FPIF, FPAF Cost Plus: Includes CPFF, CPIF, CPAF Other: Includes programs that transitioned from one contract type to another during the course of the evaluation.
Software Lifecycle	Req'ments, Design; Code; Test; Support	Early: Requirements, Design. Late: Code, Test, Support.
Language %	45% to 100%	Bi-level: 100% vs Less than 100%.
Application Type	Avionics, Command & Control, Database, Simulation, Other	Real Time: Includes Avionics, Simulation, Command and Control Information System: Includes database, other.
Budget	Budget at Completion; Latest Revised Estimate	Low: Below Average Budget. High: Above Average Budget.
Budget Volatility	Low, Medium, High	Low; Med, High: Based on Program personnel assessment.
Size	Source Lines of Code (SLOC)	Small: < 100 K LOC. Large: > 100 K LOC.
New/Modified Code	Percentage New/Modified Code	High: > 90% New/Modified code. Low:< 90% New/Modified code.
Requirements Volatility	Low, Medium, High	Low; Med, High: Based on Program personnel assessment.
Rebaselining	Yes/No	Yes/No: Based on Program personnel assessment.
Quality Standards	Yes/No	Yes: Quality standards are on contract. No: Quality standards are not on contract.
Quality Parameters	Yes/No	Yes: Quality metrics reported to program office. No: Metrics not reported to program office.
Program Activity (derived moderator)	> .01 of budget expended over 12 month period	< 0.01 of budget expended over 12 month period, the data point was excluded. ≥ 0.01 of budget expended: include data point.
Percent Complete (derived moderator)	< 80% complete > 80% complete	< 80% complete. > 80% complete.
Baseline Volatility (derived moderator)	% change in BAC over 12 month period	Low: Change in BAC < 15% during period. High: Change in BAC > 15% during period.

4.6 Summary

The results of our research design and data collection methodology provided a dataset which is sufficiently large and generally representative of the population of interest. In the following chapters, this dataset is analyzed and conclusions are drawn regarding the nature of the correlation between CMM rating and project success.

5. Analysis

5.1 Overview

Our analysis phase consisted primarily of obtaining information about the distribution of the performance indices SPI and CPI at three of the five levels of SEI CMM maturity rating. This was done in order to ascertain the nature of the correlation, if any, between performance indices and CMM rating levels. In addition, various filters and sorts were applied to the dataset to discern the effect of moderators on the SPI/CPI - rating correlation. The results of the analyses are presented as scatter plots and box & whiskers plots to show central tendency and variation. Nonparametric analysis of variance was applied to refine the analysis and to support the conclusion derived therefrom.

In this chapter, the data analysis and the results of that analysis are presented according to the hierarchy shown in figure 5-1. The analysis is performed on what we call the "complete dataset." The complete dataset, is derived from the gross dataset by purging questionable low-activity data points. The complete dataset was first evaluated *in toto*, then filtered by moderators relating to the CMM rating, and by moderators relating to cost and schedule performance. The moderators relating to CMM rating which are of greatest interest are "Rating Relevance," which relates to the associative relevance of the performance indices to the rating, and "Rating Type" which relates to the method used to obtain the rating. The moderators relating to cost and schedule performance of most interest can be loosely grouped into those moderators which relate to programmatic issues such as "Baseline Volatility," "Contract Type," and "Percent Complete," and those moderators which relate to technical issues such as "Application Type," "Programming Language," and "Project Size."

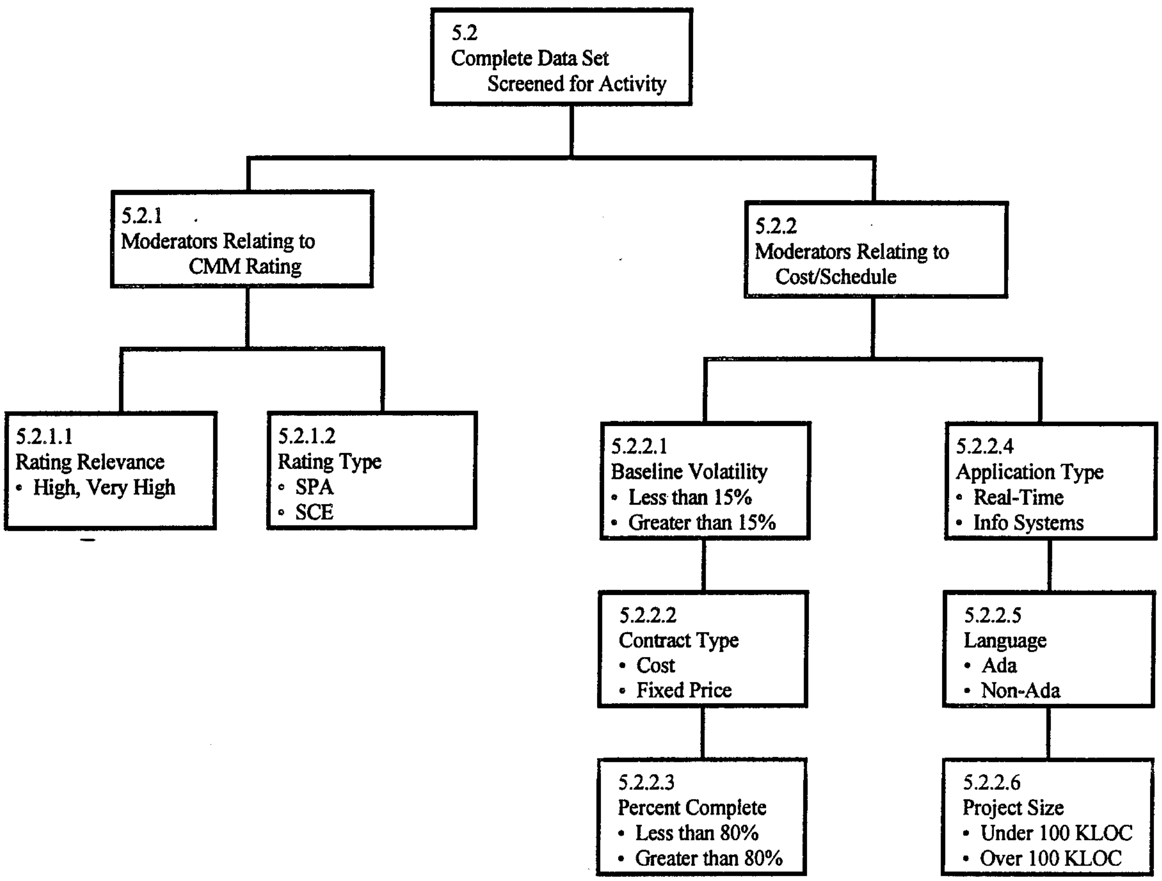

Figure 5- 1 Flow of Analysis

Analyses are presented in separate "Cases" which correspond to the Data Analysis Flow Diagram in figure 5-1. In each Case, the effect of each moderator was analyzed by filtering the complete dataset using the coding scheme developed for the moderating variable. The resulting set of data points was subjected to the following analytical tools:

1. Scatter Plot of the dataset: CPI and SPI versus Rating -- provides a means for visual inspection of the relationship between the variables.
2. Box and Whiskers plot of the dataset -- provides a pictorial summary of the datasets' more prominent features, including center, spread, extent and nature of any departure from symmetry, and any outliers (Devore, 1982:27).
3. Kruskal-Wallis nonparametric analysis of variance -- quantitatively establishes whether there is a difference in the performance index medians among the CMM ratings.
4. Multiple Comparison Test -- quantitatively establishes which of the performance index distributions associated with each rating are statistically distinct from the other distributions.
5. Descriptive statistics -- displays the mean, median, and standard deviation of the performance indices at each level of CMM rating.

In addition to the above, the following tools were applied to the complete dataset, to establish the degree of normality of the sample.

6. Histogram of the frequency density for each rating level -- provides an indication of the nature of the distribution, its central tendency, and skew.
7. Wilk-Shapiro evaluation of normality at each level -- quantitatively indicates the degree of normality of the CPI and SPI indices at each CMM rating.

5.2 *Analysis of the Complete Dataset*

Scatter Plot of the Dataset: One of the most efficient ways to get a sense of the correlation between independent and dependent variables is to create a scatter plot, where the treatments (in our case, ratings) are plotted along the abscissa, and the response (in our case, the performance indices, CPI and SPI) are plotted along the ordinate.

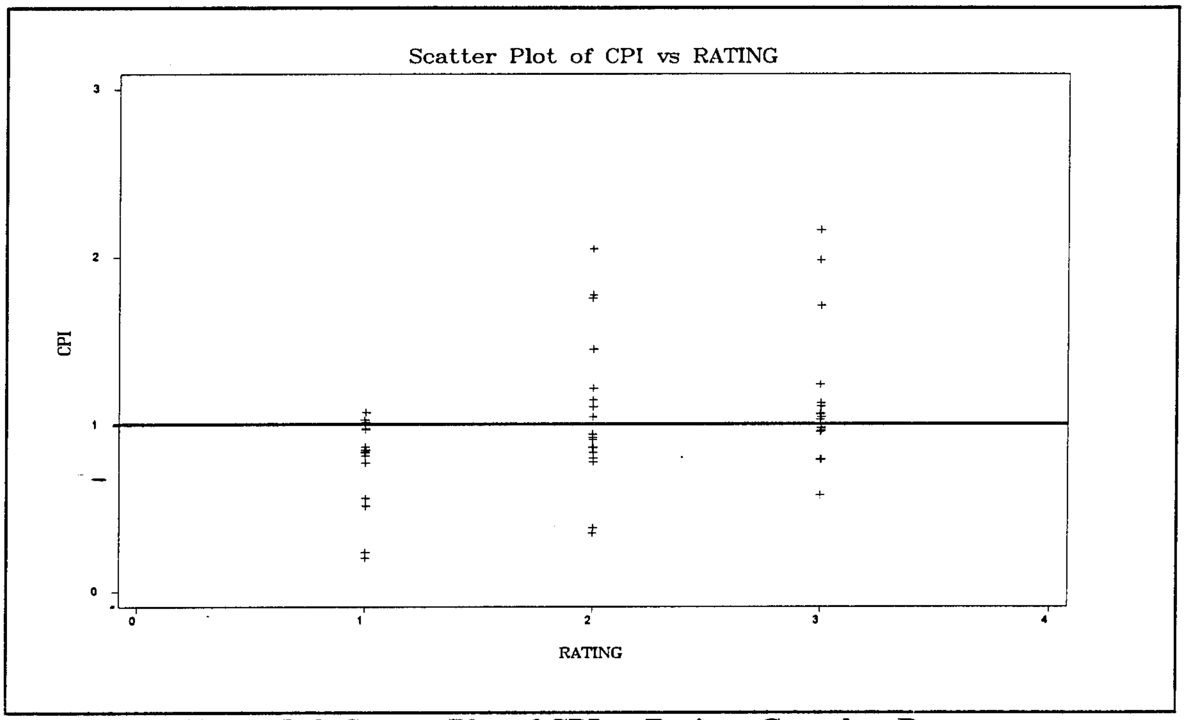

Figure 5- 2 Scatter Plot of CPI vs Rating--Complete Dataset

The characteristics that are immediately apparent about the relationship between Cost Performance Index and rating in figure 5-2 is that CPI generally increases with increasing rating. Note that the majority of Level 1 CPI data points are below a CPI of 1.00. This shows that most Level 1 projects in our dataset exhibit a cost performance generally lower than planned, resulting in a cost overrun during the 12-month period surrounding the rating. With increasing rating, the number of data points above a CPI of

1.00 increases. This suggests a trend toward improving cost performance among Level 2 and three contractors. That there appears to be a clustering of data points around a CPI of 1.00, particularly at a rating of Level 3, suggests that more "mature" contractors are able to more consistently keep their costs in line with their budgets.

The characteristic that is immediately apparent about the relationship between Schedule Performance Index and rating in figure 5-3 is the marked decrease in the variation of SPI at rating levels above Level 1. This indicates that more "mature" contractors are better able to maintain their schedules than Level 1 contractors. Also note that at rating Level 3, the number of data points above an SPI of 1.00 appears to be proportionally greater than at Level 1 or Level 2. This indicates that the most mature contractors may tend to post schedule underruns.

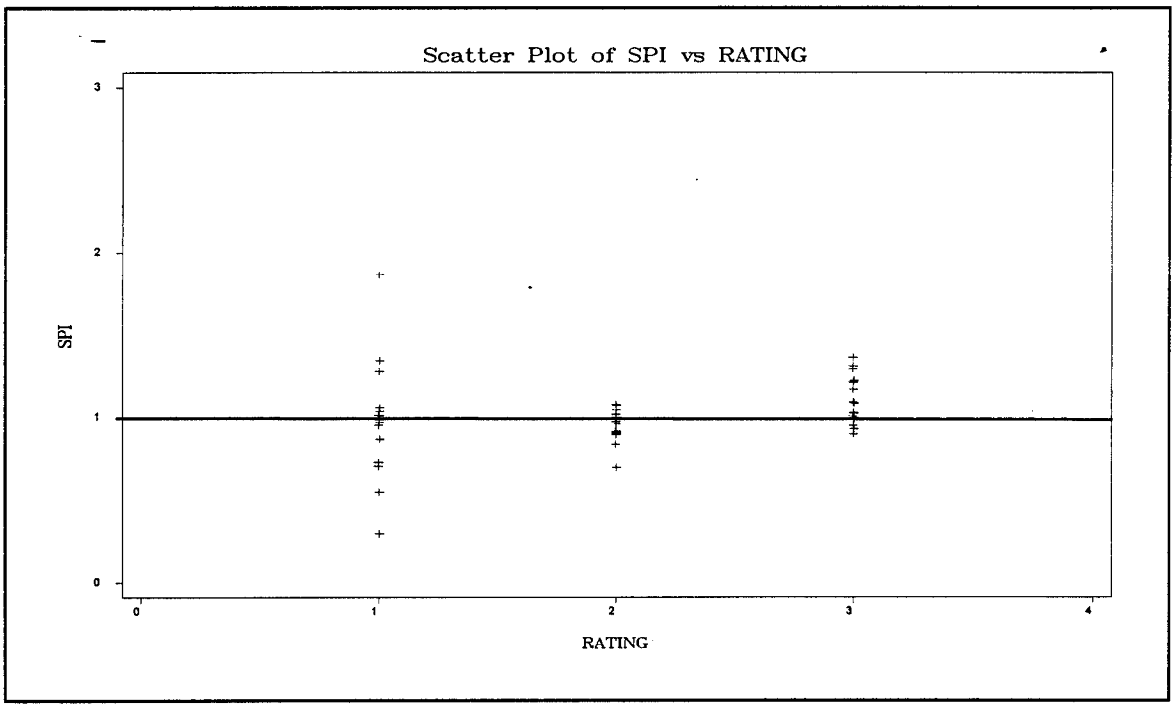

Figure 5- 3 Scatter Plot of SPI vs Rating--Complete Dataset

Histogram of the Complete Dataset: The histogram of the distribution of the performance indices at each level describes the nature of the distribution of performance indices at each rating level. The histogram indicates immediately the central tendency, the "shape" and "spread" of the data.

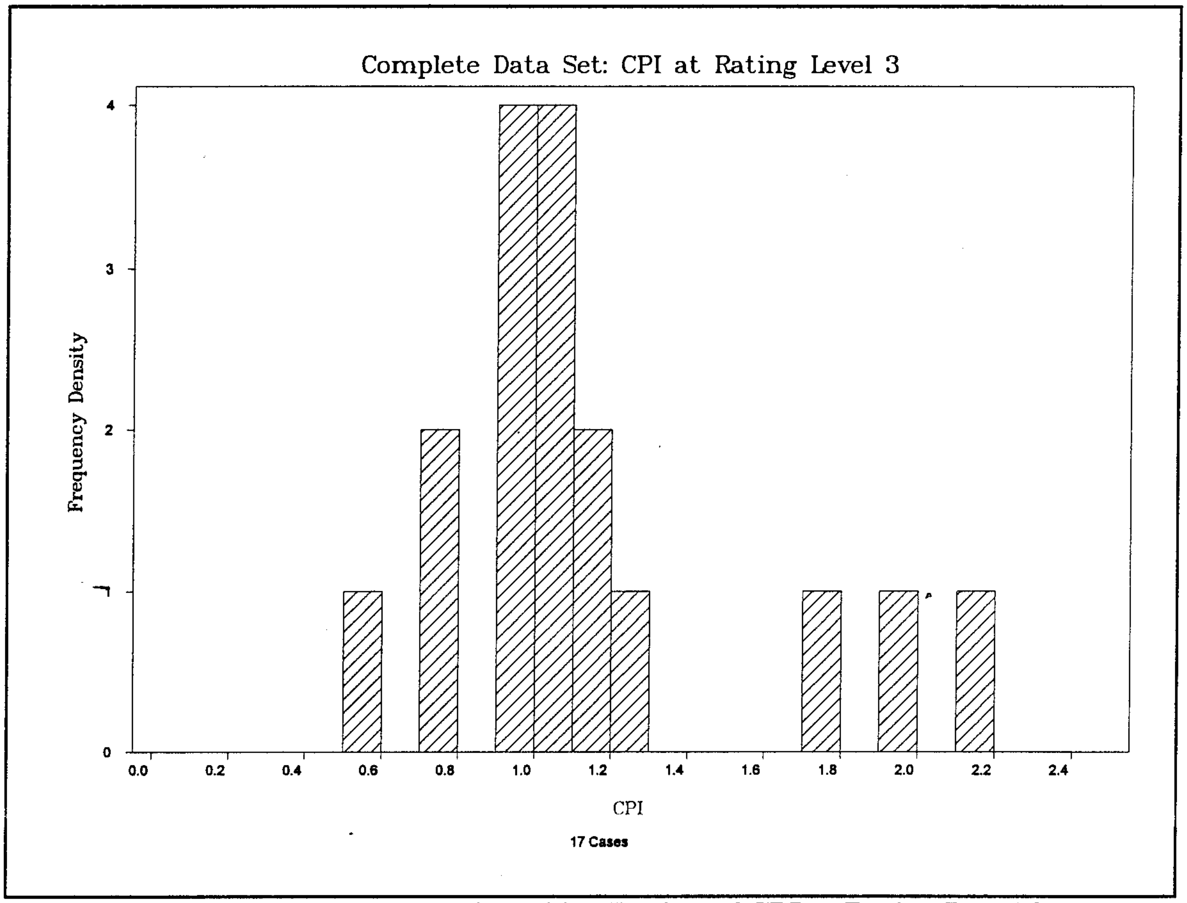

Figure 5- 4 Histogram of the Distribution of CPI at Rating Level 3

For both CPI and SPI, the performance indices at each level of CMM rating demonstrate a general "mounded shape" characteristic, similar to that shown above in figure 5-4. The shape of the distribution is significant in subsequent analyses of variance. The Kruskal-Wallis nonparametric analysis of variance assumes a chi-squared distribution, which is mound shaped, and originates at a value of zero. The histograms suggest that this assumption is not inappropriate for the dataset. The complete set of histograms showing the frequency density for each performance index at each level of CMM rating for the complete dataset is provided at Appendix C.

Wilk-Shapiro Normality Test: The Wilk-Shapiro test both visually and numerically articulates the degree to which the data approximate a normal distribution. The normal distribution and the chi-squared distributions are intimately related (Devore 1982:162). The normal distribution, as with the chi-squared distribution, are both related to the behavior of natural phenomena, and are frequently used in the analysis of categorical data, and of human and economic behaviors (Devore 1982). Table 5-1 gives a summary of the Wilk-Shapiro normality test statistic for the complete dataset. The closer the statistic comes to a value of 1.00, the more approximately normal the distribution of the performance index at the given rating level is. Customarily, distributions with Wilk-Shapiro values above 0.8 can be considered relatively normal. Given the interrelatedness of the normal and chi-squared distributions, and the Wilk-Shapiro results shown in table 5-1, the assumption of either a normal or a chi-squared distribution of the data is not inappropriate.

Table 5- 1
Wilk-Shapiro Normality Test Results for the Complete Dataset

	Rating Level 1	Rating Level 2	Rating Level 3
CPI	0.8439	0.9245	0.8105
SPI	0.8806	0.8958	0.9525

The complete set of Wilk-Shapiro plots are presented in Appendix C.

Box and Whiskers Plot of the Complete Dataset: Even more than the scatter plot, the box and whiskers plot succinctly presents important aspects of the data--particularly central tendency, spread, and outliers--enabling rapid assessment of the nature of the correlation.

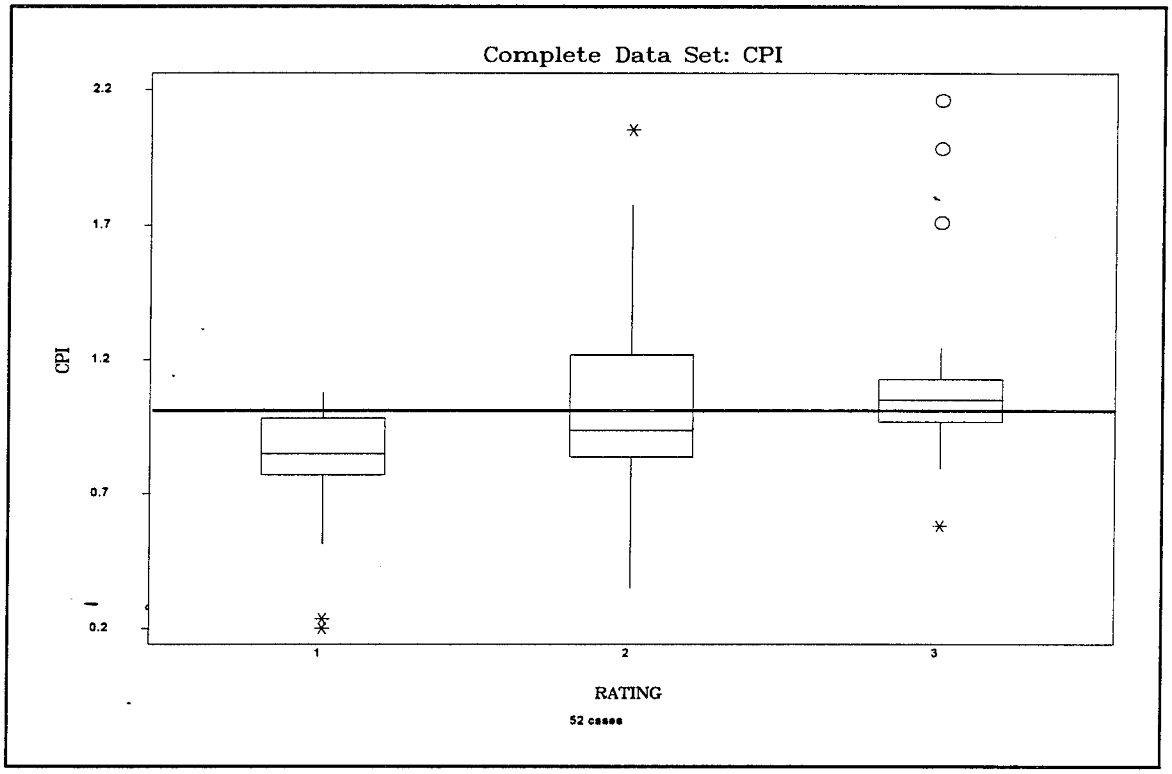

Figure 5- 5 Box and Whiskers Plot of CPI vs Rating--Complete Dataset

Figure 5-5 clearly shows the increasing central tendency of CPI with increasing rating level. The horizontal bar runs through the chart at approximately a CPI of 1.00. The box for each rating level encloses the middle half of the data points, and is bisected by a line which indicates the median of the data points. Note that the median of the Level 1 CPI is below a CPI of 1.00, and the median of Level 3 CPI is above 1.00, emphasizing the trend observed in the scatter plot. Note also the spread of the data (indicated by the length of the box) is smaller at Level 3 than at Level 1 and Level 2-- lending credence to the observations made of the scatter plot, that the most mature organizations meet their cost plans with greater certainty. The whiskers (vertical lines emanating from the ends of the boxes) indicate the range of "typical" data values--longer whiskers are indications of greater overall sample variance. The box and whisker plots also show "outliers"--extreme values in the dataset, which may be anomalies. Possible outliers are indicated by asterisks, probable outliers are indicated by circles. (Analytical Software 1992:97-98). The reader should bear in mind that outliers may significantly affect the value of statistics such as mean and variance. We attempted to mitigate the effect of outliers by using the sample median as the statistic of central tendency. The sample median is less sensitive to outliers (Devore 1991:18).

This box-and-whiskers plot of SPI versus rating (Figure 5-6) shows that the central tendency of SPI at all rating levels hovers closer to an SPI of 1.00 than did the CPI. Note also the spread of the data (indicated by the length of the boxes and whiskers) is generally narrower than that observed for CPI. The conclusion that this observation suggests is that SPI is less sensitive to rating level than is CPI.

In contrast to the observations made of the SPI scatter plots (Figure 5-3), the distinct decrease in variation of SPI from rating Level 1 to 2 and 3 is less evident in the box and whiskers plots. This decrease may indicate that the large variation observed in the SPI scatter plot for Level 1 contractors is more an effect of outliers than any significant difference in the data distributions.

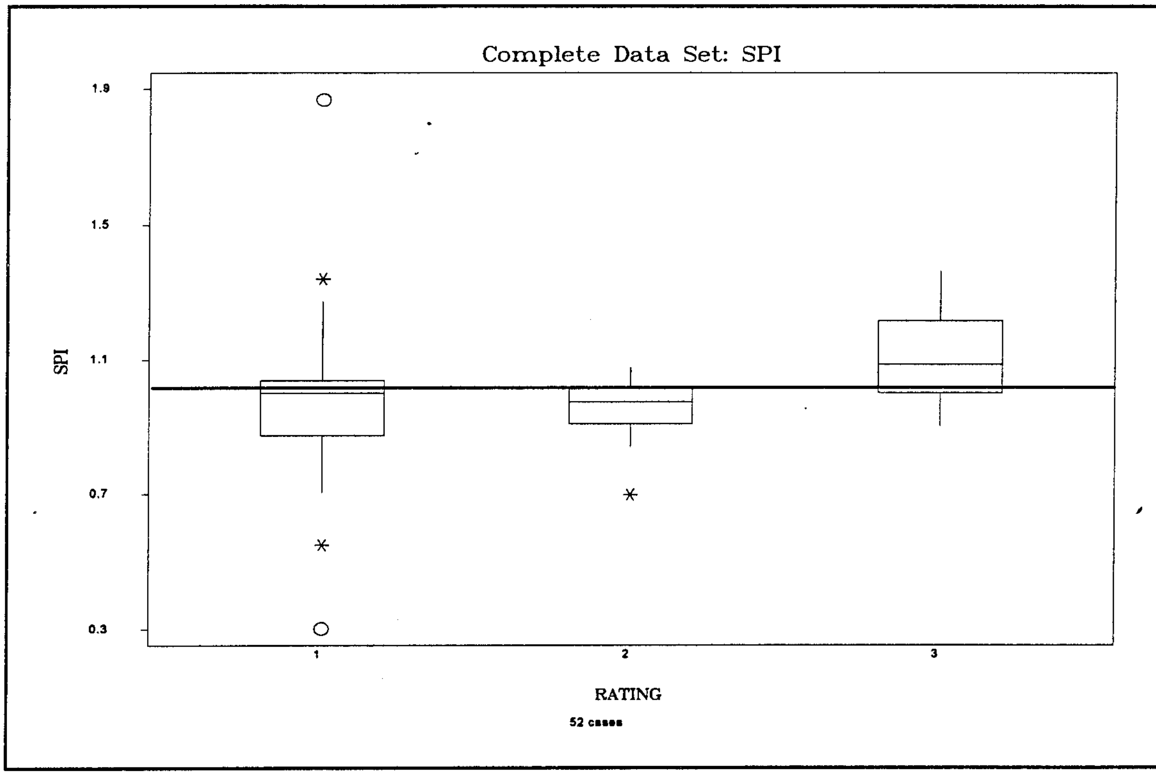

Figure 5- 6 Box and Whiskers Plot of SPI vs Rating--Complete Dataset

Kruskal-Wallis nonparametric analysis of variance: As explained in the chapter on methodology, the purpose of the Kruskal-Wallis one-way nonparametric analysis of variance is to determine if a set of data grouped by treatment (in our case, rating) is all of one distribution, or is made up of distinct distributions. The consequence of such an analysis is to determine if the various treatments (ratings) actually have a significant "effect" on the dependent variable (in our case, CPI and SPI); in which case, the different ratings will result in distinct distributions of SPI and CPI. Such a test will show if there is a significant difference in performance between, say, a Level 1 organization and a Level 3 organization. The Kruskal-Wallis test assumes a null hypothesis of no significant difference between the distributions of the three treatments (rating levels). The test then calculates the probability that this null hypothesis is correct--that there is in fact no statistically significant distinction between the distributions at the three rating levels. The P-value is the numerical expression of the probability that the null hypothesis is correct. If the P-value is below the critical value established by the confidence level of the test (in our case, for a 95% confidence level, the critical value is 0.05), then the null hypothesis must be rejected in favor of the alternate hypothesis--namely, that the distributions are actually distinct.

For the distribution of CPI for the complete dataset, the P-value of 0.016 is below the significance level of 0.05 (Table 5-2), which indicates there is a statistically significant distinction in CPI between at least two of the three rating levels. Given that there is a distinction between median CPIs of at least two of the three rating levels, a multiple comparison was run to determine which rating levels differ, and how each stacks up relative to the others.

Table 5- 2
Kruskal-Wallis Test For the Complete Dataset--CPI

KRUSKAL-WALLIS ONE-WAY NONPARAMETRIC AOV FOR CPI BY RATING

RATING	MEAN RANK	SAMPLE SIZE
1	18.3	17
2	28.2	18
3	32.9	17
TOTAL	26.5	52

KRUSKAL-WALLIS STATISTIC	8.2319
P-VALUE, USING CHI-SQUARED APPROXIMATION	0.0163

For the distribution of SPI for the complete dataset, the P-value of 0.017 is below the significance level of 0.05 (Table 5-3), which indicates there is a statistically significant distinction in CPI between at least two of the three rating levels. Given there is a distinction between the SPIs of at least two of the three rating levels, a multiple comparison was also performed.

Table 5- 3:
Kruskal-Wallis Test For the Complete Dataset: SPI

KRUSKAL-WALLIS ONE-WAY NONPARAMETRIC AOV FOR SPI BY RATING

RATING	MEAN RANK	SAMPLE SIZE
1	24.2	17
2	20.8	18
3	34.9	17
TOTAL	26.5	52

KRUSKAL-WALLIS STATISTIC	8.1238
P-VALUE, USING CHI-SQUARED APPROXIMATION	0.0172

Multiple Comparison Test: Once having established a statistically significant difference between the medians of at least two of the three groups (using Kruskal-Wallis), a test of simultaneous multiple comparison was then performed to identify which samples differed from the others. We established the direction of the difference by noting the relative magnitude of the sample mean ranks.

The multiple comparison matrix (Tables 5-4 and 5-5) displays the results of a three-way comparison between the three rating levels, articulating the significance of the differences between means of the ranks (calculated by the Kruskal-Wallis test) for each rating level. The numbers of the matrix are calculated by subtracting the absolute value of the difference between the mean of the ranks for each rating, and the right-hand-side of the multiple comparison inequality (the calculations supporting the multiple comparison matrices are provided at Appendix C).

A positive value in any matrix cell indicates there is a statistically significant distinction between the performance indices of the pair under comparison (confidence level of 80%). A negative value indicates there is no significant difference between the distributions of the performance indices of the pair of ratings. In the case where a significant difference is found, the relative magnitude of the median rank determines which rating has the greater median performance index.

In the case of the CPI for the complete dataset (Table 5-4), there is a significant distinction between the cost performance of level 1 organizations and level 2 organizations, and an even greater distinction between the cost performance of level 1 organizations and level 3 organizations. However, there is no significant distinction between the cost performance of level 2 and 3 organizations. This conclusion is intuitively consistent with the observations made of the box and whisker plots, and scatter plots above, but are lent statistical validity by the application of these simple, but powerful tests.

Table 5- 4:
. Multiple Comparison Matrix for CPI

			Rating		
Rating	n	Mean Rank	1	2	3
1	17	18.3		--	--
2	18	28.2	0.5		--
3	17	32.9	5.067	-4.7	

K-W Statistic of 8.2319, P=0.0163
Note: Shaded cells denote significant difference in sample mean ranks.

Similar to the above discussion for the multiple comparison test of the CPI data, the three-way analysis of the SPI data (Table 5-5) yields interesting conclusions about the nature of the correlation between the ratings and their respective performance indices. These tests indicate that Level 3 organizations outperform Level 1 and Level 2 organizations in terms of schedule performance.

Table 5- 5:
Multiple Comparison Matrix for SPI

			Rating		
Rating	n	Mean Rank	1	2	3
1	17	24.2		--	--
2	18	20.8	-6.0		--
3	17	34.9	1.167	4.7	

K-W Statistic of 8.1238, P=0.0172
Note: Shaded cells denote significant difference in sample mean ranks.

The combination of the Kruskal-Wallis and the multiple comparison tests confirm that there is indeed a statistically significant distinction between some levels of CMM rating and the indices of project success (CPI and SPI). By statistically significant, we mean that the Kruskal-Wallis statistic identified the difference of medians to a confidence level of 95%, and the multiple comparison test determined the relative rank to a confidence level of 80%.

Descriptive Statistics for the Complete Dataset: The Descriptive statistics for the complete dataset are provided in table 5-6 below. The statistics, when combined with the analyses above, clarify the nature of the correlation between CMM Rating and the performance indices CPI and SPI.

Table 5- 6:
Descriptive Statistics for the Complete Dataset

	Rating=1 CPI	Rating=1 SPI	Rating=2 CPI	Rating=2 SPI	Rating=3 CPI	Rating=3 SPI
N	17	17	18	18	17	17
Mean	0.7909	0.9816	1.0685	0.9562	1.1537	1.1059
Std Dev	0.2639	0.3366	0.4502	0.0915	0.4165	0.1433
Min	0.2019	0.3028	0.3496	0.6978	0.5808	0.8998
Median	0.8493	1.0000	0.9365	0.9727	1.0498	1.0864
Max	1.0788	1.8676	2.0506	1.0774	2.1602	1.3652

With regard to cost performance, we see an increasing median CPI between Level 1 and Level 3. However, the multiple comparison test indicates there is a significant distinction only between Level 1 and Level 2, and between Level 1 and 3. The distinction between the medians of Levels 2 and 3 are not significant to an 80% confidence level. The same pattern of increasing central tendency is observed with the means of the CPI for the three rating levels.

The variation of CPI data (expressed by the standard deviation) from level to level shows no trend of improvement (reduction) with increasing rating level, contrary to the scatter plot which shows a tighter grouping of the data points at Level 3. This increase in variation may be due to the presence of several outliers, as depicted in the box-and-whiskers plots.

With regard to SPI, there appears to be little difference between the means and medians at the various rating levels. This lack of apparent difference shows how the Kruskal-Wallis and multiple comparison tests can provide insight that would otherwise be absent. These tests show significant difference between Level 1 and Level 3, and between Level 2 and Level 3 SPI. These tests indicate that Level 3 organizations may outperform Level 1 and Level 2 organizations in terms of schedule performance.

5.2.1 *Analysis of Moderators Relating to CMM Rating.*

During the course of a statistical analysis, one must identify an "independent variable," and a "dependent variable." We designated the rating levels as the "independent variable," or "treatment," and the performance indices as the "dependent variable," or "response," as if this were an experiment, and we were observing the effect on the performance index as we varied the rating level. In actual fact, we were not conducting an experiment and did not have any more control over the "treatment" than we did the "response," so it is not inappropriate to discuss the factors influencing the "treatments" or the CMM rating levels, and observe how these factors may affect the correlation between the ratings and performance indices.

The first factor we suspected would have an important moderating effect had to do with the associative relevance of the performance data to the CMM ratings. We called the moderator "rating relevance." At the simplistic level, the logic goes like this: The CMM rating refers to the organization, the organization conducts the program, the project is part of the program, and the cost data describe performance on the project. Thus the rating and the performance data are mutually relevant. The flaw in this logic is that the organization may have several discrete sub-organizations, each of which may have different processes and procedures. It is conceivable that the different sub-organizations may have different levels of process maturity. The CMM rating process evaluates only a subset of all the work an organization does, and bases its conclusions in part on those sample projects. Given that not every project performed by the organization is closely associated with (and therefore representative of) the rating, it is reasonable to characterize the degree of association between the project being evaluated and the CMM rating of interest. This moderator thus helps capture the degree to which a project is representative of the maturity of an organization at the time it is rated.

The second moderator of interest is the rating method. We found, during the course of our literature review, that the different methods of determining the maturity of the software development process within an organization may result in different ratings. To explore how this dichotomy affected our dataset, we stratified our sample on the rating method: SCE versus SPA. Note that the SPA is an assessment conducted for the subject organization, with a focus toward process assessment and improvement. There were two sub-categories of SPA which we became aware of during the course of the data collection. One type of SPA, which we called "internal" was performed by the organization itself, often with specially-trained teams performing the assessment. The other type of SPA, "external," was conducted by a paid outside organization, either a contractor, or the SEI itself. These two subtypes are identified within the dataset, but were not taken into account for the following analysis. The SCE, on the other hand, is done by the government, to evaluate the suitability of the organization to perform on a contract.

5.2.1.1 Analysis of the Moderator "Rating Relevance"

The moderator "Rating Relevance" is the first of the two rating-related moderators we considered. This moderator relates to the degree of association between the project evaluated and the rating of the organization.

The striking characteristic of the CPI and SPI for high and very-high relevance data points, is that the behavior is quite similar to that which was observed for the complete dataset, except that the outliers for CPI Level 3 (Figure 5-7), and SPI Level 2 (Figure 5-8) are gone. The trend of performance indices observed for the highly-relevant dataset strengthens the observation made earlier for the complete dataset; namely, that for CPI, the Level 1 data have a high variance, and are almost exclusively below a CPI of 1.00, the Level 2 data have a high variance, but are centered on a CPI of 1.00, and the Level 3 data are centered on 1.00, but have a relatively low variance.

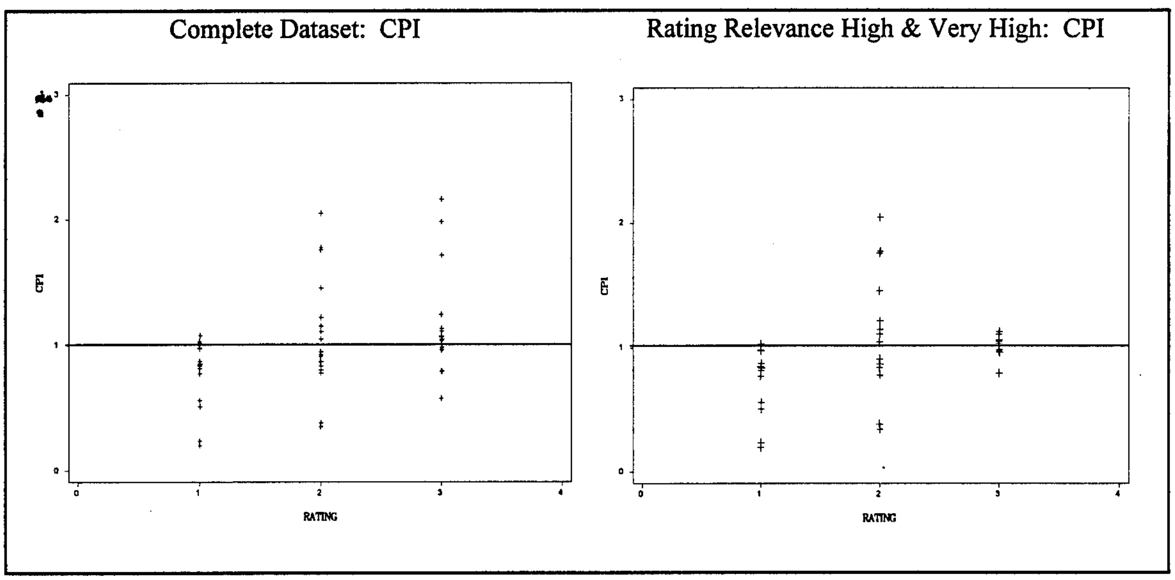

Figure 5- 7 Scatter Plots of CPI for the Complete Dataset and High & Very High Rating Relevance

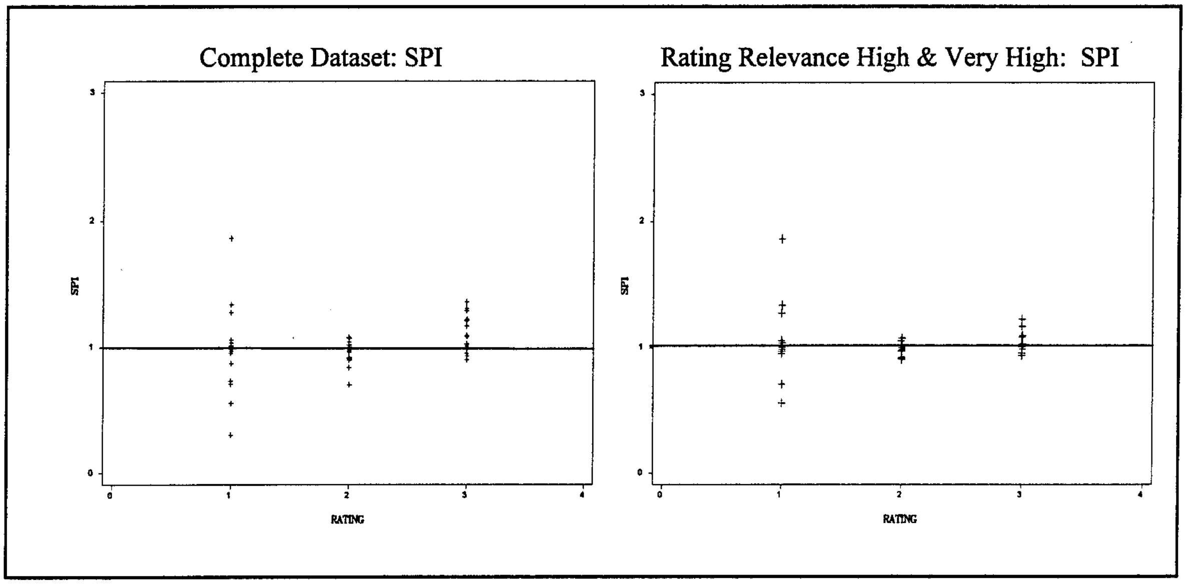

Figure 5- 8 Scatter Plots of SPI for Complete Dataset and High & Very High Rating Relevance

This phenomenon for the trend of increasing central tendency (median) and decreasing variance in CPI is vividly illustrated by the box and whisker plots of the high-relevance dataset (Figure 5-9).

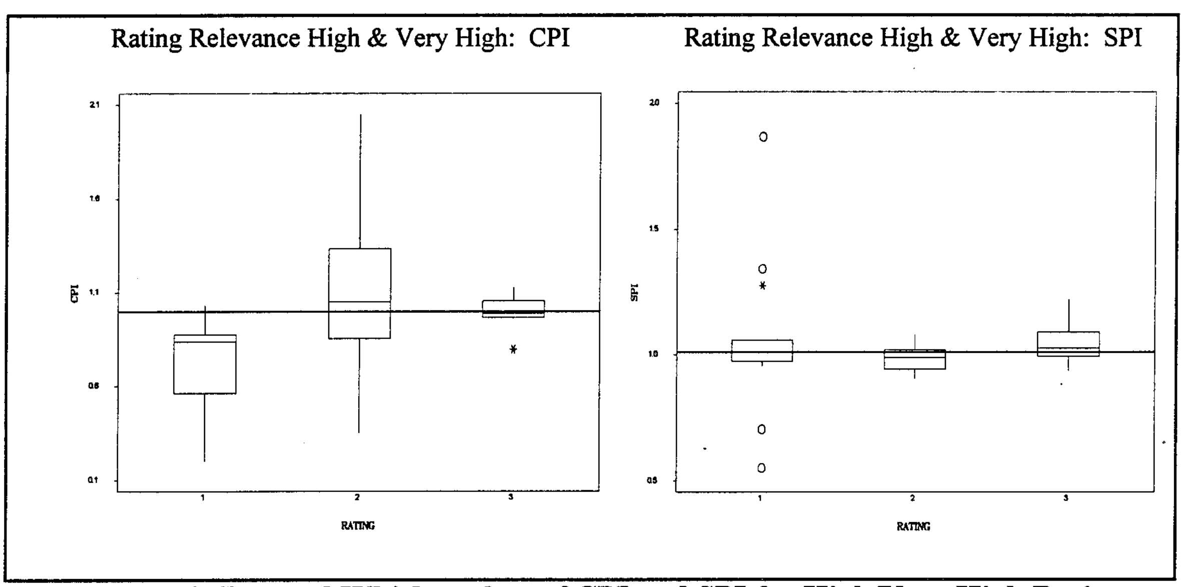

Figure 5- 9 Box and Whisker plots of CPI and SPI for High/Very High Rating Relevance

Note that SPI shows no significant trend in the data (Figure 5-9). For both the complete dataset and the high-relevance dataset, SPI tends to remain at a value of 1.00. This is borne out by the Kruskal-Wallis and multiple comparison tests, which show no statistically significant difference in the distributions of SPI at rating Level 1 through Level 3 (Table 5-7).

Table 5- 7
Multiple Comparison Matrix for SPI--High and Very High Rating Relevance

			Rating		
Rating	n	Mean Rank	1	2	3
1	14	21.5		--	--
2	15	16.8	-3.267		--
3	11	24.3	-5.839	-1.011	

K-W Statistic of 2.7738, P=0.2498

On the other hand, the Kruskal-Wallis and multiple comparison tests clearly indicate that for CPI, the high-relevance dataset shows significant distinction in the distributions for Level 1 and Level 2, and Level 1 and Level 3. As with the complete dataset, the high-relevance CPI shows no significant distinction in the distributions between Level 2 and Level 3 (Table 5-8).

Table 5- 8:
Multiple Comparison Matrix for CPI--High and Very High Rating Relevance

			Rating		
Rating	n	Mean Rank	1	2	3
1	14	13.0		--	--
2	15	24.7	3.733		--
3	11	24.4	2.761	-8.211	

K-W Statistic of 8.8692, P=0.0119

For the complete dataset, we were able to observe a trend in CPI versus rating level. This correlation showed overall decreasing variance and a sample median trend toward a CPI of 1.00 between CMM rating Level 1 and Level 3. This trend was more clearly evident for the high and very high rating relevance dataset. Recall that high and very high rating relevance means that the projects from which we collected our data had been used to obtain their organizations' rating. Thus for those projects which have the highest associative relevance to the rating, the observed trend is more firmly established.

Significantly, the apparent correlation between SPI and CMM rating observed at the complete dataset level disappeared with the high-relevance dataset. This phenomenon suggests that perhaps the initial observations indicate a stronger relationship than may actually exist. The disparity between the observed behavior of the complete dataset and the high relevance dataset begs further analysis with other moderators to identify the conditions which affect SPI performance.

The complete set of analytical plots and tables for the moderator "Rating Relevance" is at Appendix D.

5.2.1.2 Analysis of the Moderator "Rating Type"

The moderator "Rating Type" was the second rating-related moderator we considered. This moderator is of interest because of the acknowledged difference in the results of the two rating methods, SPA and SCE (Bessleman, Byrnes, Lin, Paulk and Puranik, 1993:24). The SPA, which is primarily used for self-assessment, comprises the bulk of the data we collected. The SCE, which is performed by the government in the context of a source selection comprises only 18 of our 52 total data points. Thus the statistical significance of any correlation in the SCE data may be tenuous. The SPA data for CPI appear to fall along the general trend observed for the complete dataset with regard to the decreasing variance from Level 1 to Level 3, and the central tendency converging upon CPI of 1.00 over the rating range (Figure 5-10).

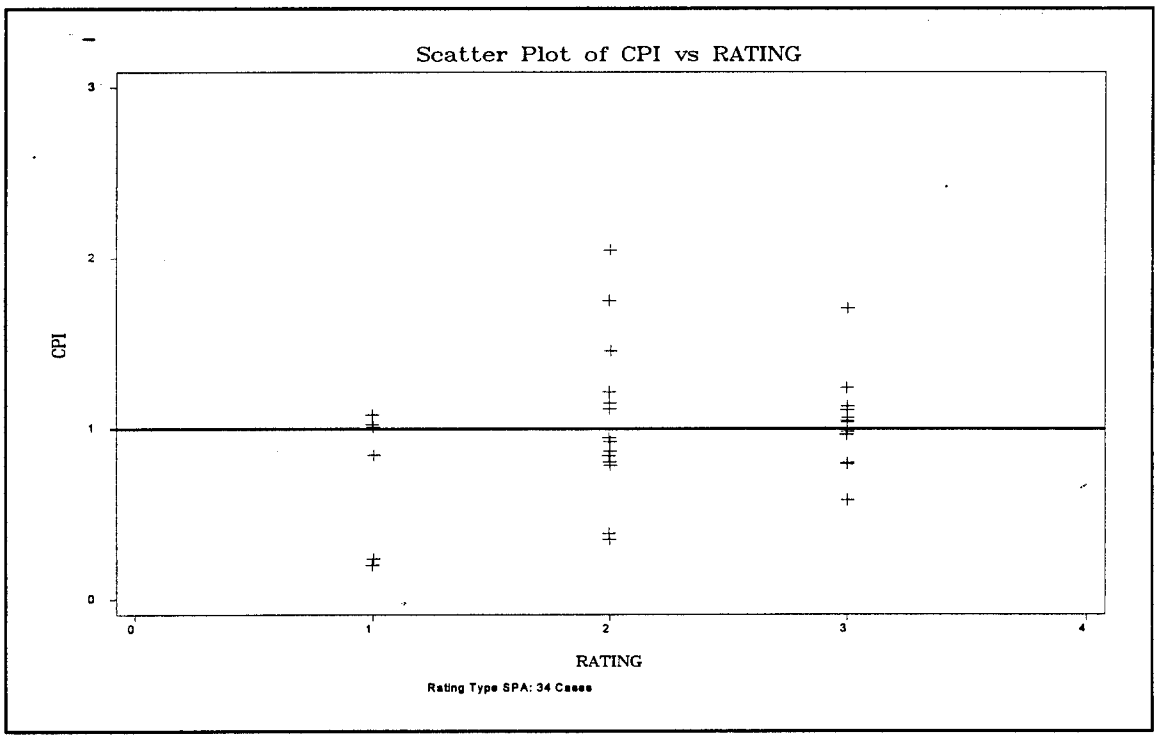

Figure 5- 10 CPI Performance of SPA Rated Organizations

However, the Kruskal-Wallis and multiple comparison tests for distinct distributions show no significant differences between any of the CPI rating distributions (Table 5-9). The lack of significant distinction between rating levels for this moderator indicates that the convergence phenomenon apparent in the plots may not be a statistically significant trend.

Table 5- 9
Multiple Comparison Matrix for CPI for SPA

			Rating		
Rating	n	Mean Rank	1	2	3
1	6	12.8		--	--
2	14	18.0	-3.712		--
3	14	19.0	-2.712	-5.903	

K-W Statistic of 1.6706, P=0.4337

Thus the analysis for the significance of rating type on the correlation between performance indices and rating levels is inconclusive. In order to improve the validity of the analysis, the sample of SCE data points must be larger, and the distribution of data points between SPA and SCE must be more balanced.

Although a gross trend between rating level and performance was not made any clearer by stratifying on rating type, some interesting observations can still be made. It is perhaps significant that of our SPA-rated data points, only 6 out of 34 (17 percent) are rated at Level 1. For our SCE-rated data points the proportion of Level 1's is 11 out of 18 (61 percent). This may reveal something about the character of the SPA versus the SCE. Specifically, the distribution of rating levels between SPA and SCE rated organizations suggests that the SPA may rate low-maturity organizations inappropriately high, and the SCE may rate high maturity organizations inappropriately low. The difference in the intent and approach to SCEs and SPAs may also contribute to this concern.

Our data do not support this concern, however. With regard to the SPA, there are indeed proportionally more Level 3 organizations. But if the SPA inappropriately over-rated these organizations, we would expect either a lower central tendency (CPI less than 1.00), or a wider variance. Such is not the case. We see that the SPA-rated Level 3 organizations are clustered around a CPI of 1.00, with little variance (Figure 5-11). This suggests that at least with regard to Level 2 and Level 3 organizations, the SPA does not inappropriately over-rate organizational maturity.

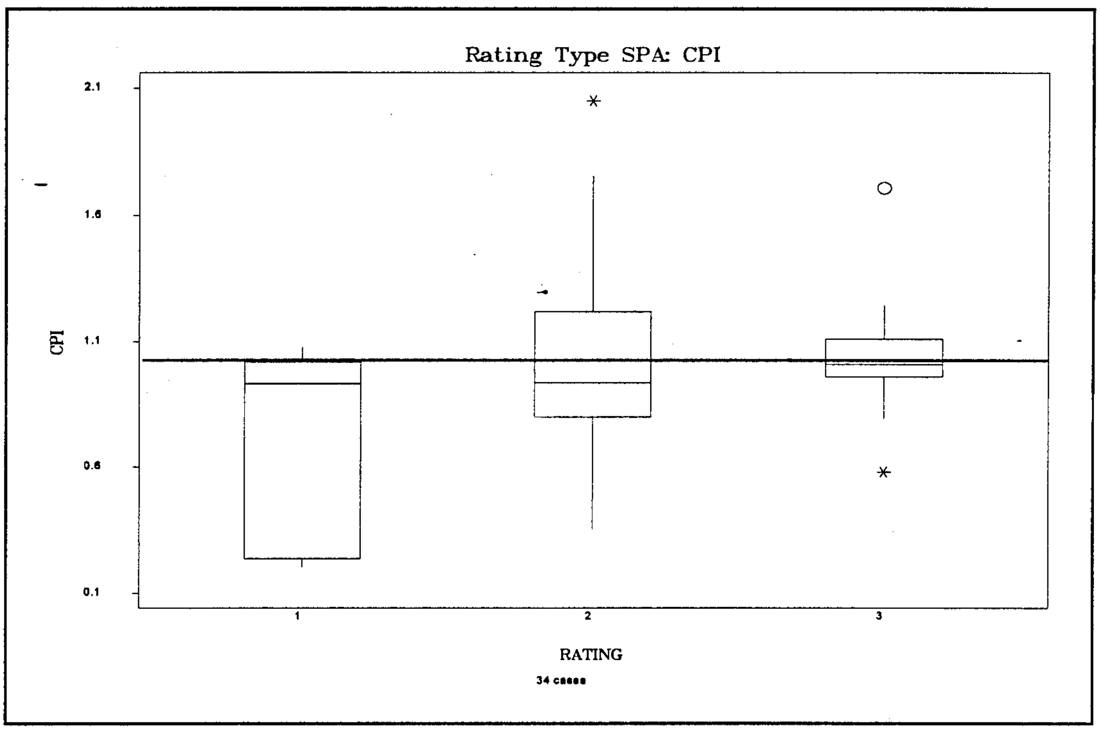

Figure 5- 11 CPI Performance of SPA-Rated Organizations

With regard to the SCE, there are proportionally more Level 1 rated organizations. If the SCE-rated Level 1 organizations were in fact Level 2 organizations, inappropriately under-rated at Level 1, we would not expect to see a trend in CPI below 1.00, as we do (Figure 5-12). This suggests that, at least with regard to Level 1 and Level 2 organizations, the SCE does not inappropriately under-rate organizational maturity.

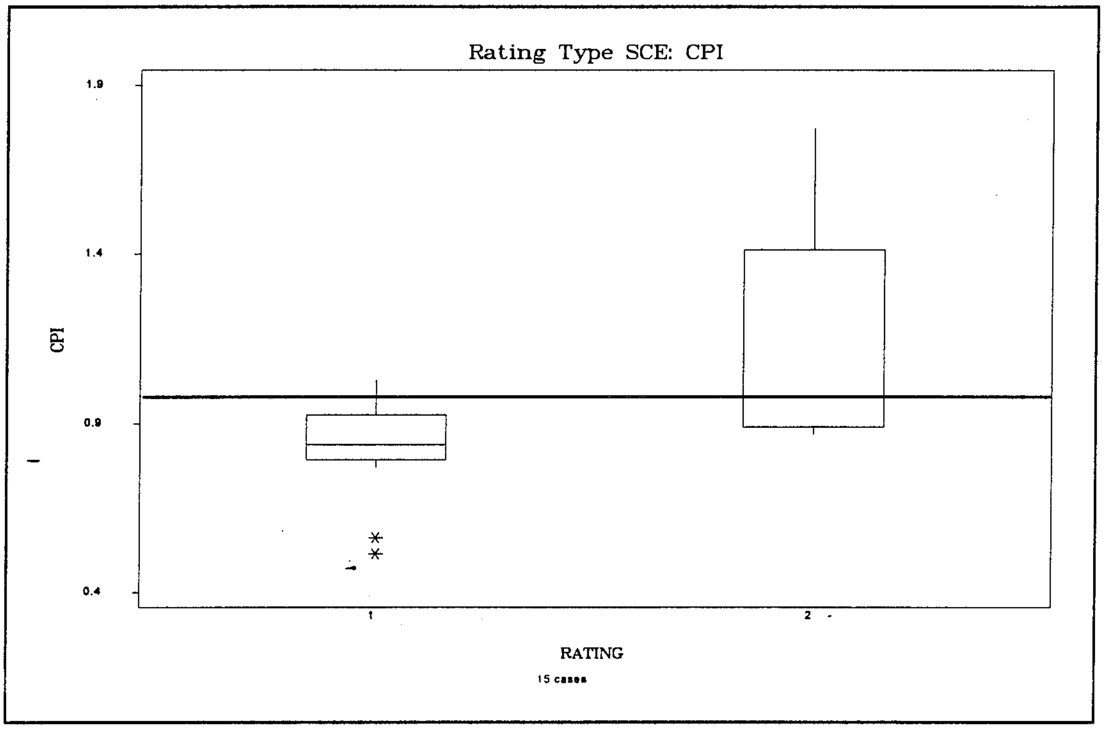

Figure 5- 12 CPI Performance of SCE-Rated Organizations

The complete set of analytical plots and tables for the moderator "Rating Type" is at Appendix E.

5.2.2 Analysis of Moderators Relating to Cost and Schedule Performance.

As the moderators relating to *CMM rating* may have influenced the nature and/or degree of the correlation between CMM rating and performance, so too may the moderators of *cost and schedule performance* have influenced the nature and/or degree of correlation between rating and performance. Conceptually, we have distinguished between those moderators which are of *programmatic* significance and those of more *technical* significance. Factors of a programmatic nature, such as "Baseline Volatility," "Contract Type," and "Percent Complete" reflect the structure of the Government/contractor relationship and the forces that act upon that relationship, as the program progresses through the acquisition cycle. The technical moderators, such as "Application Type," "Language," and "Project Size" attempt to capture the essential qualities of program size and complexity, which may influence the overall difficulty of the program, and thus the contractor's success in its execution.

5.2.2.1 Analysis of the Moderator "Baseline Volatility"

A correlation between rating and performance could be affected by the relative changes in the baseline of a project. These changes in baseline can take many forms: an increase/decrease in the scope of work (Engineering Change Proposals, Technical Change Proposals, etc.), transfer of tasks from one WBS element to another, reallocation of management reserve, or a formal reprogramming (negotiating an over-target baseline). The causes of these changes in baseline can vary from redirection on behalf of the government to inadequate initial budgeting by the contractor. It is possible that a change in the baseline, regardless of the type, could affect the link between contractor performance and the performance indices of interest. In addition to concerns of cost growth and schedule delay triggered by such changes, the Government is concerned that any rebaseline may provide an opportunity for the contractor to obscure unfavorable cost and schedule variances from the baseline (Christensen 1994).

To address this concern, we examined the proportional change in the budget-at-complete (BAC) over the period of interest, i.e., we calculated the change in total budget over the 12 month period as a percentage of the budget at the beginning of the period. This rate of change of the budget is indicative of rebaselining, whatever the source, whether it is due to reallocation of work, ECPs, or reprogramming. We arbitrarily selected a change in budget of plus or minus fifteen percent as the stratification level in our analysis.

The effect of this moderator is significant in that programs which show a high degree of baseline volatility exhibit no statistically significant difference in cost and schedule performance, whereas programs which show a relatively low degree of baseline volatility demonstrate the same general increase in performance as was observed in the complete dataset case. This distinct difference in performance trends between these two levels of baseline volatility are clearly seen in figures 5-13 and 5-14.

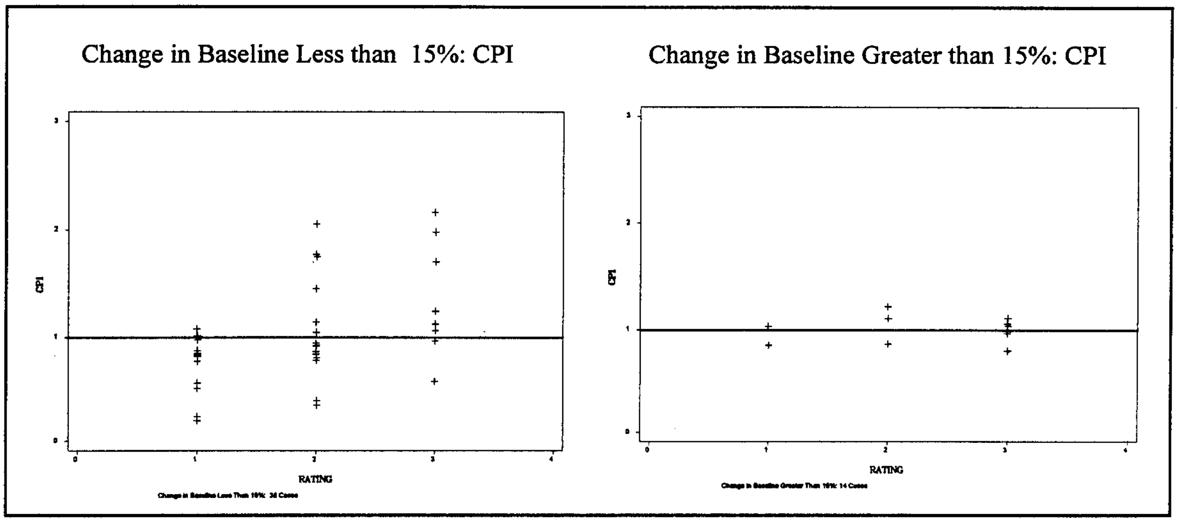

Figure 5- 13 Comparison of CPI trends for moderator "Baseline Volatility"

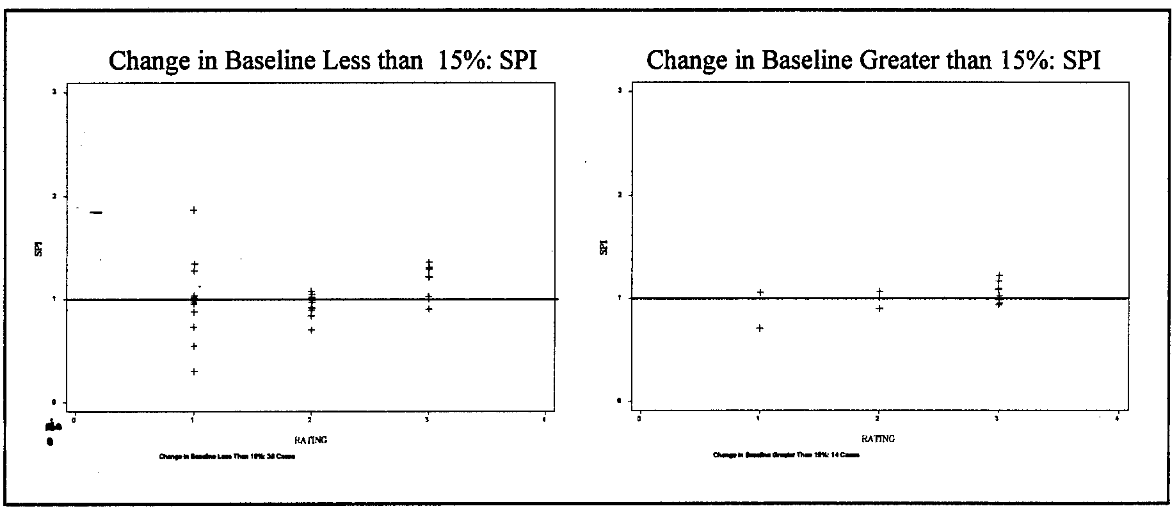

Figure 5- 14 Comparison of SPI trends for moderator "Baseline Volatility"

Substantiating the observation of distinct performance trends are the results of the Kruskal-Wallis and multiple comparison tests. In the case of those programs with baseline volatility less than fifteen percent, there is a statistically significant difference in the distributions of Level 1 and Level 3 cost and schedule performance indices.

However, for those programs exhibiting a baseline volatility greater than fifteen percent, there is no statistically significant difference between any of the levels (Appendix F). The complete set of analytical plots and tables for the moderator "Baseline Volatility" is at Appendix F.

5.2.2.2 Analysis of the Moderator "Contract Type"

The type of contract used to procure systems fundamentally influences the relationship between the Government and the contractor. For example, a fixed-price contract tends to place the monetary risk on the contractor, while a cost-type contract shifts most of the monetary risk to the Government (Nicholas 1990:497). The apportionment of risk between the parties affects how the task is proposed, costed, structured, performed, and tracked. Such a profound environmental moderator may have an effect on the correlation between performance and rating.

Though the scatter plots of the data show no obvious distinction between the cost-type and fixed-price type contracts (Appendix G), the descriptive statistics (Appendix G) appear to show a consistently higher mean CPI for fixed-price contracts at each level than for cost contracts. In other words, the fixed-type contracts show a CPI trend which is "shifted upward" in comparison to the cost-type CPI data (Table 5-10). This "shift" may be due to the fact that on a fixed-price contract, the contractor increases profit when it underruns the cost baseline. This upward-shift in performance for fixed-price contracts is also observed for SPI, and (except for level 1) is as prominent as with CPI. It is possible the fixed price contracts provided incentives for beating the baseline schedule, which might account for the shift, however our dataset did not include this level of detail. Thus it is not evident from the data that contract incentives were the cause of the shift.

Table 5- 10:
Comparison of Mean Performance for Cost-type and Fixed-type Contracts

		Rating=1 CPI	Rating=1 SPI	Rating=2 CPI	Rating=2 SPI	Rating=3 CPI	Rating=3 SPI
Cost Contracts	N	3	3	5	5	9	9
	Mean	0.6191	0.9608	0.9316	0.8905	1.1001	1.0811
Fixed-Price Contracts	N	4	4	9	9	8	8
	Mean	0.8284	0.9433	1.1081	0.9611	1.2139	1.1338

The complete set of analytical plots and tables for the moderator "Contract Type" is at Appendix G.

5.2.2.3 Analysis of the Moderator "Percent Complete"

In our review of the literature, we found that proximity to completion has a significant effect on the dynamics of the cumulative performance indices. For example, cumulative SPI, by definition, is driven to 1.00 at program completion while cumulative CPI has been shown to be stable from the 20% completion point, where "stability" is defined as CPI range being less than 0.2. The dynamics of the cumulative performance indices have been well noted in the literature, and are a fundamental element in the art of estimating at-complete costs (Christensen and Heise 1993:7-15) In our research, however, we are taking a 12-month slice of these performance indices. We acknowledge these "snapshot" indices will not be as stable as the cumulative indices. Nevertheless, it was important that our research capture the degree to which the dynamics of the cumulative indices affected our non-cumulative indices.

For SPI, as stated above, the nature of the index is such that at program completion, it is identically equal to 1.00. That is, at the completion of the contract, all budgeted work packages are complete, and earned value has been taken. For contractors which have fallen behind schedule during the course of the contract, and have demonstrated a cumulative SPI below 1.00, one would expect disproportionately high non-cumulative SPI over the latter stages of contract performance, in order for cumulative SPI to equal 1.00 at contract completion. In our analysis, we define program percent complete as the percentage of earned value taken relative to total budget. Therefore, the concept of program completion is not linked to a chronological schedule (i.e., completion date), but to the amount of work done relative to the amount of work required by the contract.

For our dataset, we chose 80 percent complete as the point about which we stratified the sample. This was done to distinguish between the performance over the bulk of the contract and the performance near program completion. As it turned out, nearly 60 percent of our sample is composed of contracts within the latter 20 percent of contract performance. Thus, any dynamics related to the latter stages of contract performance may affect the correlation between the ratings and the performance for the overall sample. For example, if non-cumulative SPI is artificially biased upward at the latter stages of contract performance, perhaps that effect swamps any maturity-related effect on SPI that may have been observable in a more representative sample.

This hypothesis is given credence by the SPI scatter and box/whisker plots (Figures 5-15 and 5-16), respectively, which for contracts less than 80 percent complete, show a maturity-related trend not unlike that observed for CPI in the overall dataset--whereas for contracts over 80 percent complete that trend is practically reversed. At less than 80 percent complete Level 1 projects are almost all below a SPI of 1.00 while at greater than 80 percent complete Level 1 projects are almost all above a SPI of 1.00. This is in contrast to SPI at rating Levels 2 and 3, which appear to remain relatively stable over the course of the contract.

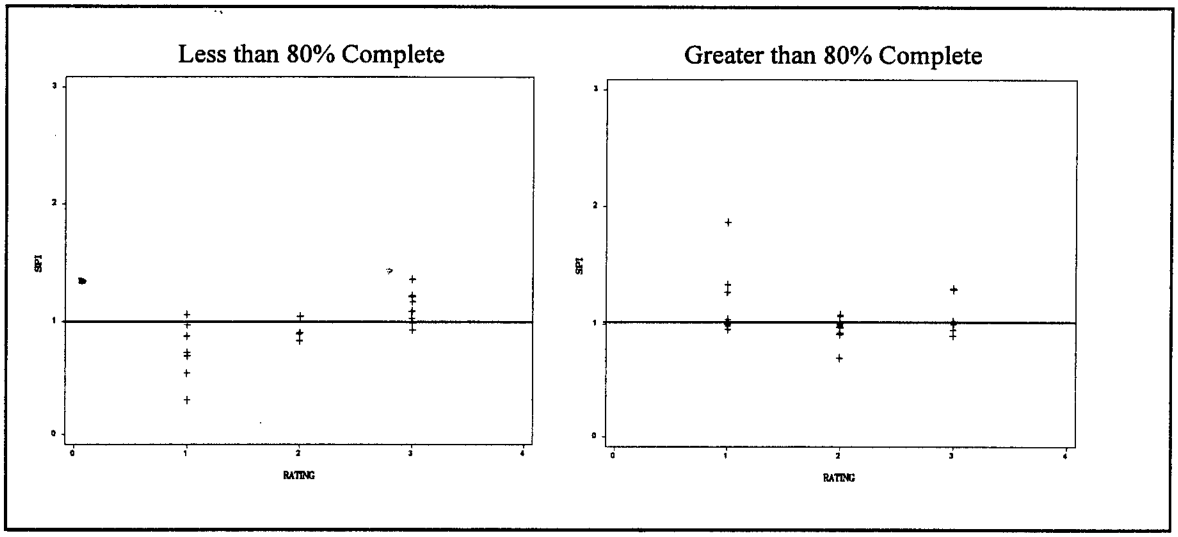

Figure 5- 15 Scatter Plots of SPI versus Rating for Contract Percent Complete

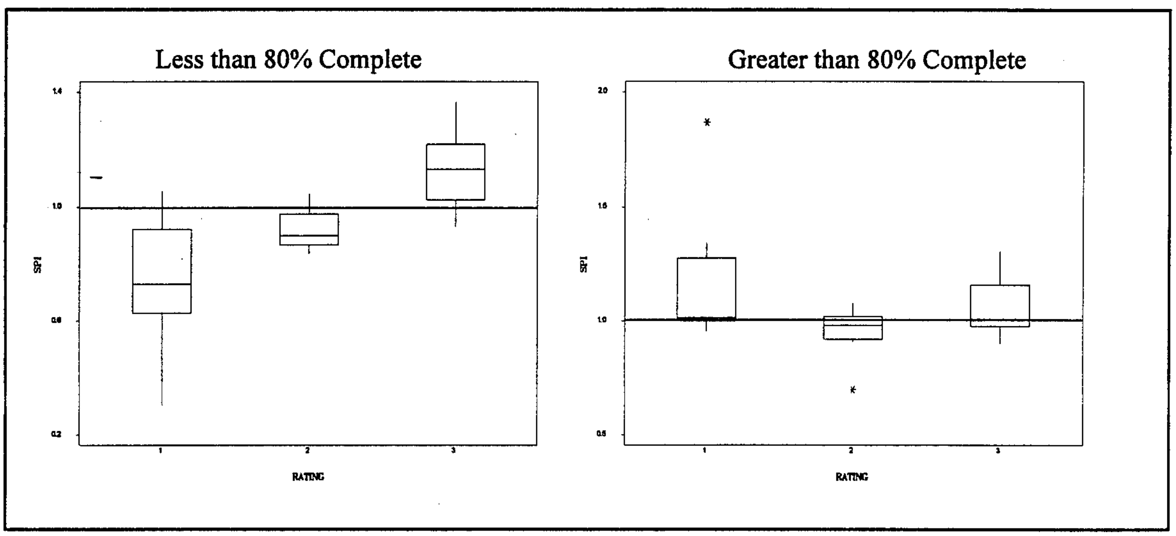

Figure 5- 16 Box & Whisker Plots of SPI versus Rating for Contract Percent Complete

With regard to CPI, the data suggest that the performance of less-mature contractors tends to be worse in the last 20 percent of contract performance, while the performance of more mature contractors tends to be better in the last 20 percent of contract performance (Figure 5-17).

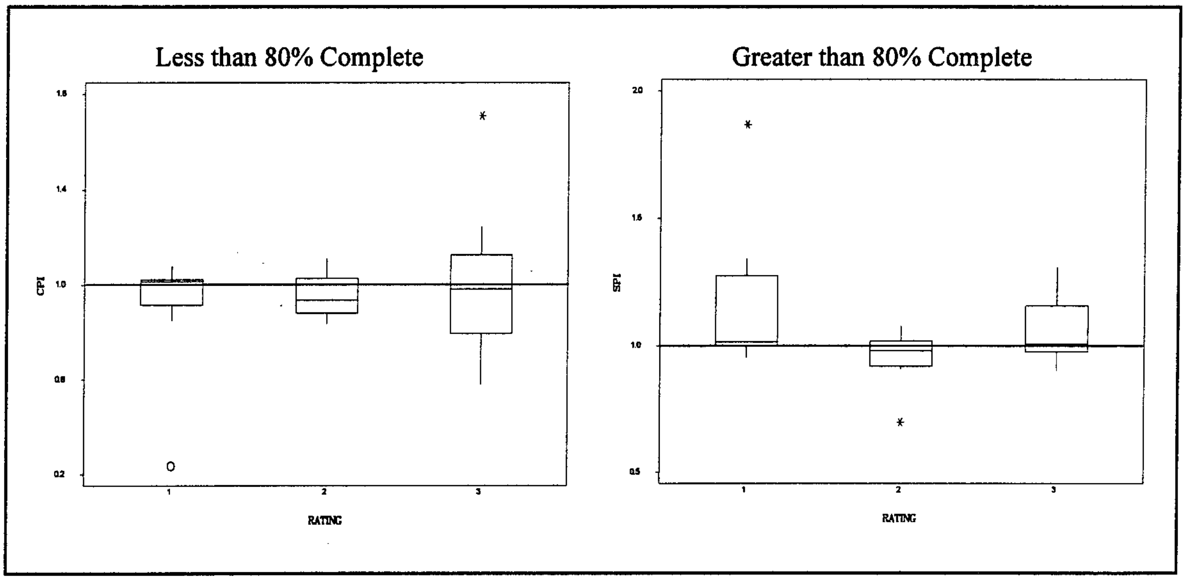

Figure 5- 17 Box & Whisker Plots of SPI versus Rating for Contract Percent Complete

The complete set of analytical plots and tables for the moderator "Percent Complete" is at Appendix H.

5.2.2.4 Analysis of the Moderator "Application Type"

Application type is a gross predictor of project complexity. The categories selected, real-time applications versus information systems applications, capture the distinction between the highly complex avionics, flight control, simulation, and command and control applications and the usually less-demanding database and catalog applications.

Our dataset shows that of the real-time applications, nearly half (12 out of 25) are associated with Level 3 contractors. The cost performance of these projects are distinctly above a CPI of 1.00, with a mean of 1.259. This is in contrast to the performance of less mature contractors, who implement real-time applications with mean CPIs of 0.77 (Level 2), and 0.72 (Level 1). The difference in performance at these levels is shown in the scatter plot of CPI versus Rating (Fig 5-18), and is substantiated by the multiple comparison test (Table 5-11). These results would suggest that the more complex applications are being implemented with apparent success by mature software development organizations.

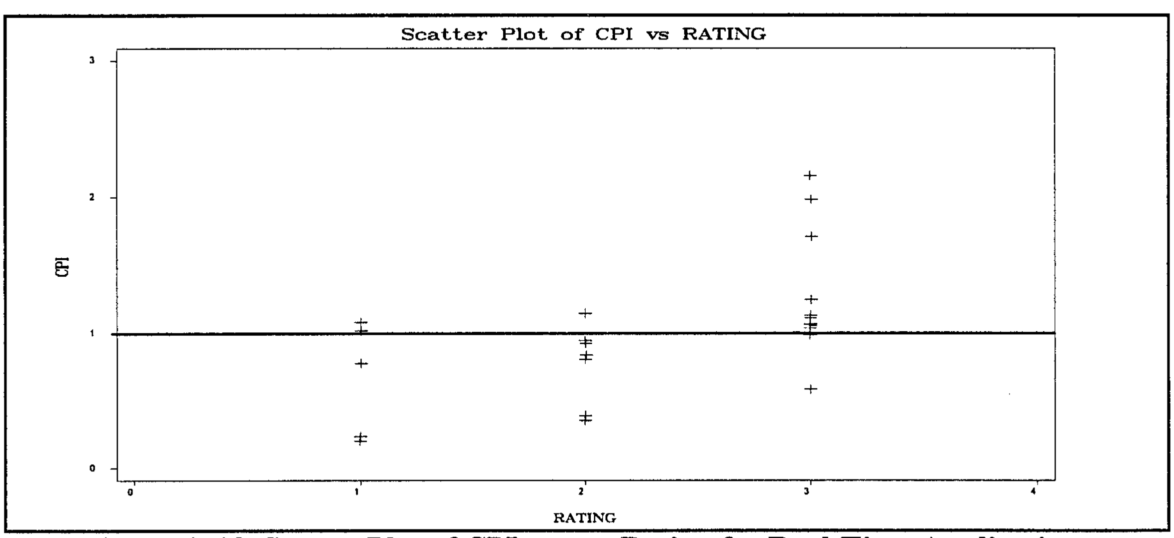

Figure 5- 18 Scatter Plot of CPI versus Rating for Real-Time Applications

Table 5- 11
Multiple Comparison Matrix for CPI for Real-Time Applications

			Rating		
Rating	n	Mean Rank	1	2	3
1	6	8.7		--	--
2	7	8.9	-7.31		--
3	12	17.9	2.451	2.58	

K-W Statistic of 8.9519, P=0.0114

In contrast to the real-time applications, only 5 of the 26 information systems projects are implemented by Level 3 organizations. The remainder are approximately evenly distributed between Level 1 and Level 2. For information systems applications, the data suggest that increased maturity does translate into substantially better cost performance from Level 1 to Level 2, but the variation of the Level 2 data is high relative to Level 1 data (Figure 5-19). The scarcity of data points at Level 3 precludes definitive analysis of the performance at that level. The complete set of analytical plots and tables for moderator "Application Type" is provided at Appendix I.

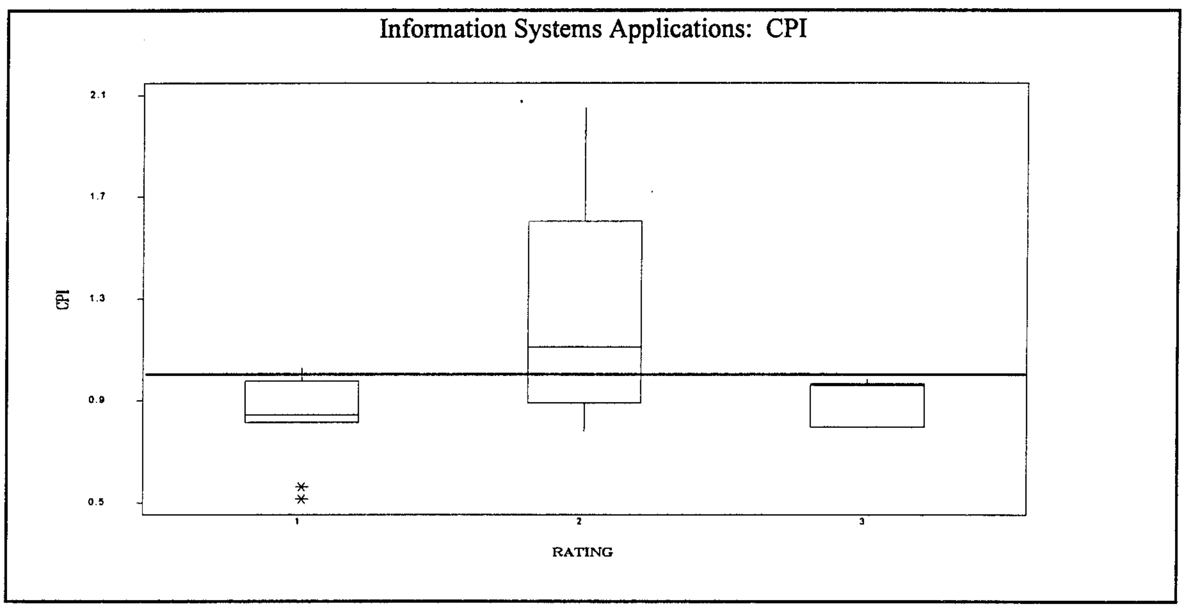

Figure 5- 19 Box Whisker Plot of CPI vs Rating for Information Systems Applications

5.2.2.5 Analysis of the Moderator "Language"

Ada, as the official "standard" higher order language (HOL) of the DoD, is mandated for all new software development programs. This requirement to use Ada may impose difficulties on software development contractors if they have little experience with Ada, or if Ada is not their preferred language. On the other hand Ada is a powerful language which imposes rigorous discipline in the development process, and thus may provide benefits in the testing and integration phases of development. Thus it is important to determine if such a significant program characteristic has any effect on the correlation between rating and performance.

The general trend of the cost performance indices with respect to rating for Ada applications is not unlike the trend observed for those applications of a real-time type. Specifically, the less mature organizations show CPI levels below 1.00 (mean CPI for Level 1 is 0.727, for Level 2 is 0.765), the Level 3 organizations have a mean CPI of 1.038. This similarity between Ada applications and real-time applications is not surprising, given that the majority of the real-time applications in our dataset are coded in Ada.

In comparing the performance between Ada and Non-Ada applications, we found that Level 1 and Level 2 organizations' mean CPIs and SPIs are lower using Ada than with languages other than Ada --the numbers show the same effect for Level 3 organizations, but the non-Ada sample size is too small for meaningful comparison (Table 5-12). Note, no test for significance was performed on the Ada/Non-Ada mean performance indices, so the reader is cautioned not to infer a statistically significant performance difference between Ada and Non-Ada projects at Levels 1 and 2.

Table 5- 12
Comparison of Mean Performance for Ada and Non-Ada Applications

		Rating=1 CPI	Rating=1 SPI	Rating=2 CPI	Rating=2 SPI	Rating=3 CPI	Rating=3 SPI
Ada Applications	N	8	8	4	4	12	12
	Mean	0.7270	0.9501	0.7648	0.8496	1.0375	1.1012
Non-Ada Applications	N	6	6	10	10	3	3
	Mean	0.8224	1.0175	1.2126	0.9823	1.7365	1.1664

This negative impact of Ada on the performance of less-mature organizations is perhaps due to the structured nature of Ada, which, in turn, demands structure of the organization. As a result, Ada may work better for those organizations with more mature processes. Less mature organizations may find that the discipline required to program in Ada imposes rigor that is incompatible with their chaotic software development paradigm.

The complete set of analytical plots and tables for the moderator "Language" is provided at Appendix J.

5.2.2.6 Analysis of the Moderator "Size"

Project size is the key driver in nearly all software cost estimation models, including REVIC (Revised Enhanced Version of Intermediate COCOMO), SEER-Software Estimation Model, and PRICE-S. Thus, project size is a necessary moderator to evaluate, in terms of its effect upon the rating/performance correlation. Given the lack of uniformity in the definition of software project size (we gathered data in the form of KSLOC, DSI, Equivalent DSI, and DSI converted from bytes), we can at best only give approximate size distinctions. Thus, we chose to stratify our sample on the relatively common size categories: "Greater than 100K LOC" and "Less than 100K LOC." This level of distinction is fairly common in the literature when distinguishing between relatively large programs and relatively small programs. As stated above, with the questionable consistency of our size data, any finer distinction would be misleading. Projects which have no size associated with them, such as management or testing WBSs, were excluded from the analysis.

For programs generally smaller than 100K LOC, the trend in the data is consistent with the trend observed for the complete dataset (Figure 5-20).

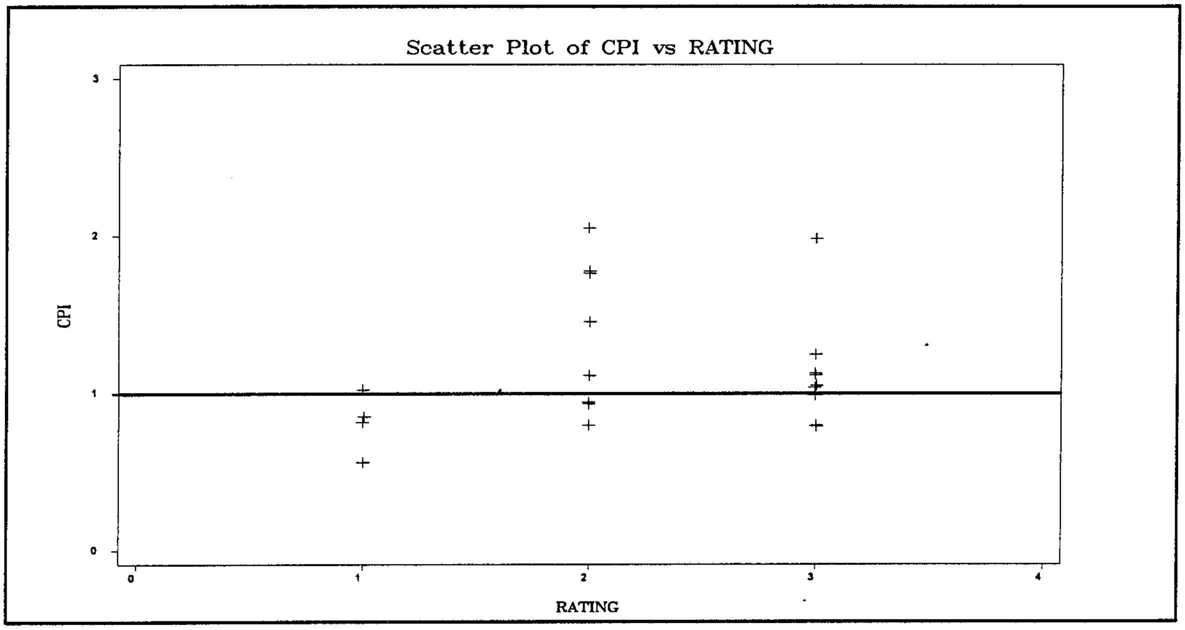

Figure 5- 20 Scatter Plot of CPI versus Rating for Applications Less than 100K LOC

However, unlike the correlation observed with the complete dataset, there is a statistically significant distinction between CPI at Level 1 and Level 2 only. This lack of distinction between CPI at Level 3 and CPI at the other levels may be a result of the smaller sample size for the moderated data. For programs greater than 100K LOC, our data show no statistically significant correlation between rating level and CPI or SPI (Appendix K).

The effect of application size on mean CPI varies with rating level. For Level 1 and 2 organizations, the larger programs tend to have lower mean CPIs than the smaller programs (Table 5-13). The size of the application does not appear to have an effect on the mean CPI for Level 3 organizations, suggesting larger applications tax the abilities of less mature organizations to a greater extent than they tax the abilities of more mature organizations.

Table 5- 13
Comparison of Mean CPI for Applications Less than and Greater than 100K LOC

		Rating=1 CPI	Rating=2 CPI	Rating=3 CPI
Applications < 100 K LOC	N	4	8	9
	Mean	0.8113	1.3524	1.1245
Applications > 100 K LOC	N	8	4	5
	Mean	0.6875	0.6801	1.1659

It is interesting to note that the more mature organizations in our dataset are developing the smaller programs (9 out of 21 projects with fewer than 100K LOC (43%) are developed by Level 3 organizations, while only 5 out of 17 projects with more than 100K LOC (29%) are developed by Level 3 organizations). This preponderance of small projects associated with mature organizations may be driven by complexity. In the case of avionics or flight controls, the smaller programs can be the most complex, and thus may represent challenging software development programs for mature contractors.

The complete set of analytical plots and tables for the moderator "Size" is at Appendix K.

5.3 Summary

The analysis of the dataset yielded interesting insights into the nature and existence of correlation between rating level and performance. Table 5-14 summarizes the results of the Kruskal-Wallis and multiple comparison tests. These tests, powerful though they are, are neither necessary nor sufficient for the existence or absence of a correlation to be declared. Instead, they provide a degree of insight into the dataset not available solely through graphical and qualitative analysis.

Table 5- 14
Summary of Nonparametric Analysis of Variance Results

Analysis Case	Significant Difference in Levels?	Number of Different Pairs	Significant Difference in Levels?	Number of Different Pairs
	CPI	CPI	SPI	SPI
Complete Dataset	Yes	2	Yes	2
High and Very High Rating Relevance	Yes	2	No	-
Rating Type - SPA	No	-	Yes	2
Rating Type - SCE	Yes	1	No	-
Baseline Volatility - Less than 15%	Yes	1	Yes	2
Baseline Volatility - Greater than 15%	No	-	No	-
Contract Type - Cost	Yes	1	No	-
Contract Type - Fixed Price	No	-	Yes	2
Percent Complete - less than 80% Complete	No	-	Yes	2
Percent Complete - greater than 80% Complete	Yes	2	* Yes	1
Application Type - Real-time	Yes	2	No	-
Application Type - Information System	Yes	1	Yes	2
Language - Ada	Yes	2	Yes	1
Language - Non-Ada	Yes	2	Yes	1
Size - less than 100K LOC	Yes	1	Yes	1
Size - greater than 100K LOC	No	-	No	-

Higher rating levels have higher mean rank of performance unless otherwise specified
* Rating level 1 showed higher SPI performance than level 2

6. Conclusion

6.1 Overview

The purpose of our research was to determine the nature of the correlation, if any, between an organization's CMM rating and the success of the organization's software development efforts. Consequently, the conclusions derived from our research should be discussed in terms of both the existence and the nature of the correlation. Moderating variables which aided in the identification and description of any relationships between rating level and performance are incorporated into the discussion. Finally, we recommend further useful work in this area.

6.2 The Existence and Nature of Correlation Between Rating and Performance

Our research leads us to conclude that a correlation exists between performance and software process maturity. We observed improved cost and schedule performance with increasing process maturity. Specifically, the least mature organizations were likely to have difficulty adhering to cost and schedule baselines. In contrast, the more mature organizations were likely to have on-baseline cost and schedule performance. We also observed that certain moderators strongly affected this correlation.

In terms of identifying the existence of a correlation, our null hypothesis was that there was no correlation. If this hypothesis were correct, we would expect to have observed no discernible trend in either the central tendency, or the variation of the sample from rating Level 1 through Level 3.

The results of our evaluation have compelled us to reject the null hypothesis in favor of the alternate hypothesis; that there is a correlation between CMM rating and performance, as represented by CPI and SPI. This conclusion is reached by the confluence of qualitative (graphical) analysis and nonparametric statistical techniques.

Although the complete dataset provided the initial indications of the correlation, the striking correlation appeared only when several moderating conditions were applied to the dataset. Specifically, the correlation between rating level and CPI was more clear with the "Rating Relevance" moderator accounted for. The correlation between rating level and SPI was evident only with the "Percent Complete" moderator accounted for.

6.2.1 The Nature of the Correlation between Rating and Cost Performance

For the complete dataset, we saw the first hint of a trend in the central tendency of CPI, specifically between Level 1 and Level 3, where the median performance increased from a CPI below 1.00, to a CPI at or very near 1.00. However, we observed no significant change in variance of CPI between Level 1 and Level 3.

When we applied the moderator "Rating Relevance," (which establishes the associative relevance between an organization's rating and the project from which the cost/schedule data were collected) the correlation between rating level and CPI became very evident. We observed trends both in central tendency and variation across the rating levels. The trend observed was high variation with central tendency below a CPI of 1.00

for Level 1; high variation and central tendency near a CPI of 1.00 for Level 2; low variation and central tendency near a CPI of 1.00 for Level 3. Additionally, the multiple comparison test showed significant distributions between Levels 1 and 2, and between Levels 1 and 3. Thus, the trend in CPI with increasing organization maturity is a CPI generally approaching 1.00, with generally decreasing variation.

6.2.2 The Nature of the Correlation between Rating and Schedule Performance

Within the complete dataset, the variation in schedule performance appears fairly constant between Level 2 and Level 3, and is markedly less than the variation in SPI at Level 1. Thus, a trend in variation with rating level is shown only between Level 1 organizations and the rest. It appears that once an organization matures beyond Level 1, variation in SPI is relatively insensitive to maturity.

Unlike the trend observed in variation, we observed no clear trend in the central tendency of SPI within the complete data set. At all rating levels, the SPI remains close to 1.00. However, when the moderator "Percent Complete" was taken into account, an intriguing correlation between rating level and central tendency of SPI manifested itself.

We noted that for projects less than 80% complete, the performance of Level 1 organizations was consistently below a SPI of 1.00. For projects greater than 80% complete, this Level 1 behavior was reversed--Level 1 organizations posted SPIs generally greater than 1.00. For Level 2 and Level 3 organizations, we observed little change in the central tendency of SPI with increasing rating level for both "young" projects (less than 80% complete), and "old" projects (greater than 80% complete).

The Level 1 SPI behavior for "young" versus "old" projects may explain why the behavior of SPI relative to rating level was not apparent for the complete dataset or other moderator groups. For Level 1 organizations, most projects less than 80% complete exhibited a SPI under 1.00; most projects greater than 80% complete exhibited a SPI over 1.00. These two groups offset each other, such that the complete data set showed the median SPI of exactly 1.00. That the "young" Level 1 projects showed schedule overruns (SPI less than 1.00), and the "old" Level 1 projects showed schedule underruns (SPI greater than 1.00) has more to do with the way SPI is calculated than any performance improvement in these organizations. Specifically, by definition, cumulative SPI is forced to 1.00 at contract completion. In order for programs which fall behind schedule early in the contract (cumulative SPI less than 1.00) to achieve this, the non-cumulative SPI late in the contract is forced above 1.00.

In other words, we observed a similar converging behavior in SPI as we did in CPI, but only in projects which are less than 80% complete--before the nature of cumulative SPI "artificially" increased performance at the end of the projects. This effect masked the central tendency behavior of SPI in the overall data set, and is the result of the sample having disproportionately more "old" projects than "young" ones.

In summary, our research leads us to conclude that a CMM Level 1 contractor is likely to have difficulty adhering to cost and schedule baselines. In contrast, a CMM Level 3 contractor is likely to have on-baseline cost and schedule performance. We also conclude that certain moderators strongly affect the observed correlation.

6.3 Recommendations for Further Research

Further analysis of the database developed for this research should be performed. We were able to examine only eight of the moderators collected/derived. We found that other means of "slicing" the data provided interesting and valuable insight into the relationships at work in the complex process the data represents. With further examination, this database, limited and flawed though it is, will reveal more knowledge about the process of software acquisition and the maturity of software development organizations.

Further work should be done to broaden the database. Of the shortcomings of our research, the most significant has to do with the representativeness of our sample. Our sample was biased toward programs at the end of program completion. We feel sure this had the effect of hiding the behavior of the SPI with respect to rating level, and may have had other effects we were unaware of. An effort should be made to collect more data from organizations rated by the SCE method. Of course, as more organizations achieve higher levels of CMM maturity, they should be added to the database. Additionally, the distribution of data points between ESC and ASC may have introduced unintended bias. Further work should be performed to incorporate data from other product centers or relevant organizations.

As the database grows, there may evolve a statistically significant sample of organizations rated multiple times, such that longitudinal studies may be performed. It would be illuminating to track SPI and CPI over time as programs achieve higher levels of maturity. Additionally, a larger database would enable simultaneous application of multiple moderators. This is not feasible for the dataset as it currently exists. Multiple applications of moderators tend to reduce the number of data points below the number where meaningful conclusions can be drawn.

The depth of the database could also be improved. Each data point could itself be the subject of an intensive case study. Our superficial treatment of moderators could only grossly characterize the dynamics peculiar to the project. If each project were to be studied in-depth, more discerning moderators could be obtained, as could more complete data for the existing moderators.

Finally, it would be valuable for future researchers to attempt to fit a distribution to the data, to develop a predictive model for contract performance based on rating level. The software development community at large may well be interested in the probability and confidence level of a certain CPI and SPI outcome given a rating level.

6.4 Conclusion

The aim of our research was to determine the nature of a correlation between the CMM rating and software development success. Though success is difficult to measure directly, by using the surrogates of cost and schedule performance, we were able to show correlation between CMM rating and the cost and schedule performance of a generally representative sample of historical software development contracts. If we were to apply this knowledge to current software development programs, we see that the CMM rating is a useful means of assessing the general likelihood of a contractor meeting the contract cost and schedule baselines.

Appendix A: Data Collection Form

This appendix contains and example of the data collection form used to guide the collection of data from the cost libraries and the program personnel interviews. The data collection form was designed so that program identification information could be disassociated from the rest of the data to ensure anonymity of the data source. After the data collection forms were completed, the program identification information was separated and secured. Only the researchers and their faculty advisors have access to the correlation matrix which links these programs to their data points.

DATA COLLECTION FORM

*********WARNING*********

- **THIS INFORMATION IS PRIVILEGED, ACADEMIC RESEARCH DATA. INFORMATION CONTAINED ON THIS COVER SHEET, AND ASSOCIATED DATA CANNOT BE RELEASED PUBLICLY WITHOUT EXPRESS WRITTEN CONSENT OF THE RESEARCHERS.**
- **THIS COVER SHEET MUST BE DISASSOCIATED WITH RELATED DATA PRIOR TO PUBLIC RELEASE.**
- **THIS DOCUMENT AND ASSOCIATED DATA CONTAIN NO CLASSIFIED, PROPRIETARY, OR CONFIDENTIAL MATERIAL.**

DATA IDENTIFICATION TAG:__________

ORGANIZATION NAME: LOCATION: MAIL CODE:

POC NAME(S): PHONE: EMAIL:

CONTRACTOR NAME:

DIVISION: LOCATION:

PROJECT NAME:

CONTRACT NUMBER: DATA ACCESSION NUMBER:

WBS INFORMATION:

WBS 1--LEVEL:	WBS NUMBER	WBS TITLE:
WBS 2--LEVEL:	WBS NUMBER:	WBS TITLE:
WBS 3--LEVEL:	WBS NUMBER:	WBS TITLE:
WBS 4--LEVEL:	WBS NUMBER:	WBS TITLE:
WBS 5--LEVEL:	WBS NUMBER:	WBS TITLE:
WBS 6--LEVEL:	WBS NUMBER:	WBS TITLE:
WBS 7--LEVEL:	WBS NUMBER:	WBS TITLE:
WBS 8--LEVEL:	WBS NUMBER:	WBS TITLE:
WBS 9--LEVEL:	WBS NUMBER:	WBS TITLE:
WBS 10--LEVEL:	WBS NUMBER:	WBS TITLE:

COMMENTS:

DATA IDENTIFICATION TAG:__________

INSTRUCTIONS: DO NOT PUT PROGRAM-UNIQUE OR CONTRACT IDENTIFICATION DATA ON THIS FORM. THE LINKAGE TO THE PROGRAM MUST BE MAINTAINED SOLELY THROUGH THE DATA IDENTIFICATION TAG.

WBS DESCRIPTIONS:

WBS 1 DESCRIPTION:

WBS 2 DESCRIPTION:

WBS 3 DESCRIPTION:

WBS 4 DESCRIPTION:

WBS 5 DESCRIPTION:

WBS 6 DESCRIPTION:

WBS 7 DESCRIPTION:

WBS 8 DESCRIPTION:

WBS 9 DESCRIPTION:

WBS 10 DESCRIPTION:

CMM DATA:

FIRST RATING: DATE OF RATING: RATING METHOD: RATING RELEVANCE:

COMMENTS:

SECOND RATING: DATE OF RATING: RATING METHOD: RATING RELEVANCE:

COMMENTS:

THIRD RATING: DATE OF RATING: RATING METHOD: RATING RELEVANCE:

COMMENTS:

DATA IDENTIFICATION TAG:_________

<u>INSTRUCTIONS:</u> DO NOT PUT PROGRAM-UNIQUE OR CONTRACT IDENTIFICATION DATA ON THIS FORM. THE LINKAGE TO THE PROGRAM MUST BE MAINTAINED SOLELY THROUGH THE DATA IDENTIFICATION TAG.

<u>MODERATING VARIABLE INFORMATION</u>

<u>PROGRAM:</u>

ACQUISITION PHASE: ACAT LEVEL:

CONTRACT TYPE:

COMMENTS:

<u>PROJECT:</u>

LANGUAGE, LANGUAGE PERCENTAGE:

SOFTWARE LIFECYCLE POINT:

APPLICATION TYPE:

BUDGET:

BUDGET VOLATILITY:

PROJECT SIZE (LINES OF CODE, FUNCTION POINTS, ETC):

NEW vs REENGINEERED:

REQUIREMENTS VOLATILITY:

PROJECT TEAM INTEGRATION (PRIME/SUB):

OTHER SIGNIFICANT MODERATING PROJECT CONDITIONS:

DATA IDENTIFICATION TAG:__________

INSTRUCTIONS: DO NOT PUT PROGRAM-UNIQUE OR CONTRACT IDENTIFICATION DATA ON THIS FORM. THE LINKAGE TO THE PROGRAM MUST BE MAINTAINED SOLELY THROUGH THE DATA IDENTIFICATION TAG.

MODERATING VARIABLE INFORMATION

COST DATA:

REBASELINING:

COST REPORTING ANOMALIES:

OTHER SIGNIFICANT MODERATING COST CONDITIONS:

QUALITY DATA:

QUALITY STANDARDS ON CONTRACT

QUALITY PARAMETERS TRACKED

OTHER SIGNIFICANT MODERATING QUALITY CONDITIONS:

PROGRAM MANAGER COMMENTS:

DATA IDENTIFICATION TAG:_________

INSTRUCTIONS: DO NOT PUT PROGRAM-UNIQUE OR CONTRACT IDENTIFICATION DATA ON THIS FORM. THE LINKAGE TO THE PROGRAM MUST BE MAINTAINED SOLELY THROUGH THE DATA IDENTIFICATION TAG.

WBS 1:

-6 MONTH
CUMULATIVE
BCWS:
BCWP:
ACWP:
BUDGET:

RATING

+/- 6 MONTH
ΔBCWS=
ΔBCWP=
ΔACWP=
ΔBUDGET=
CPI=
SPI=

+6 MONTH
CUMULATIVE
BCWS:
BCWP:
ACWP:
BUDGET:

-3 MONTH
CUMULATIVE
BCWS:
BCWP:
ACWP:
BUDGET:

+/- 3 MONTH
ΔBCWS=
ΔBCWP=
ΔACWP=
ΔBUDGET=
CPI=
SPI=

+3 MONTH
CUMULATIVE
BCWS:
BCWP:
ACWP:
BUDGET:

WBS 2:

-6 MONTH
CUMULATIVE
BCWS:
BCWP:
ACWP:
BUDGET:

RATING

+/- 6 MONTH
ΔBCWS=
ΔBCWP=
ΔACWP=
ΔBUDGET=
CPI=
SPI=

+6 MONTH
CUMULATIVE
BCWS:
BCWP:
ACWP:
BUDGET:

-3 MONTH
CUMULATIVE
BCWS:
BCWP:
ACWP:
BUDGET:

+/- 3 MONTH
ΔBCWS=
ΔBCWP=
ΔACWP=
ΔBUDGET=
CPI=
SPI=

+3 MONTH
CUMULATIVE
BCWS:
BCWP:
ACWP:
BUDGET:

Appendix B: Gross Dataset

This appendix provides the gross dataset which was derived from the data collection forms. This database was constructed using Microsoft Access version 2.0. The database was constructed in a flat file format, with each database record representing an individual data point, comprised of identifying code, rating information, cost/schedule information, and moderating characteristics. The data is presented in a "form" format, with each record (data point) represented by a separate page. Each field in the form corresponds to a field in the database with the exception of the dependent variables and derived moderators which were calculated from the dataset.

Gross Data Set

Record ID: A A

Data Identification

Program Tag: A RatingTag: A Project Tag (WBS#): 1

Project Description: Operational mission software planning, requirememts analysis, change review/assessment, review/approval requirements specifications

Rating Information

Rating Date: 10/15/93 Rating: 3 Rating Type: SPA (EXT) Rating Relevance: Med

Rating Comment:

Moderating Variables

Acquisition Phase: Support/Upgrade Contract Type: CPI

Program Comments:

S/W Lifecycle: Requirements Language: Ada Language %: 100.00%

Application: Avionics Project Budget: 16608000 Budget Volatility: Low

Size: 156800 % New/Modified Code: 100.00% Requirements Volatility: Unk

Rebaselining : No Quality Stds On Contract: X Quality Params Tracked : X

Cost Accounting Anomalies: Variances may be influenced by letter contract prior to periods of interest

Program Manager Comments: Size was converted from bytes to DSI

Cost Data

	Six Months Prior to Rating	Three Months Prior to Rating	Three Months After Rating	Six Months After Rating
Date:	5/30/93	8/30/93	1/30/94	4/30/94
BCWS:	3110	3612	4427	4635
BCWP:	2715	3139	4040	4797
ACWP:	4096	5313	5827	7681
Budget:	16782	16782	16633	16608
LRE:	16031	16168	15541	16698

Derived Moderators

Budget Volatility Index: -0.0104 LRE Volatility Index: 0.0416 Percent Complete: 0.2888

BCWS Activity: 0.32902 BCWP Activity: 0.43402 ACWP Activity: 0.46674

Dependent Variables

Schedule Performance Index: 1.365246 Cost Performance Index: 0.58075

Investigator Comments:

Data Identification

Program Tag: A **RatingTag:** A **Project Tag (WBS#):** 2

Project Description: Planning and integration of operational mission software

Rating Information

Rating Date: 10/15/93 **Rating:** 3 **Rating Type:** SPA (EXT) **Rating Relevance:** Med

Rating Comment:

Moderating Variables

Acquisition Phase: Support/Upgrade **Contract Type:** CPI

Program Comments:

S/W Lifecycle: Integration **Language:** Ada **Language %:** 100.00%

Application: Avionics **Project Budget:** 5186000 **Budget Volatility:** Low

Size: 0 **% New/Modified Code:** 0.00% **Requirements Volatility:** Unk

Rebaselining : Yes **Quality Stds On Contract:** ☒ **Quality Params Tracked :** ☒

Cost Accounting Anomalies: Variances may be influenced by letter contract prior to periods of interest--check for rebaselining

Program Manager Comments: BCWS decreased

Cost Data

	Six Months Prior to Rating	Three Months Prior to Rating	Three Months After Rating	Six Months After Rating
Date:	5/30/93	8/30/93	1/30/94	4/30/94
BCWS:	191	365	318	365
BCWP:	155	239	320	367
ACWP:	101	160	186	225
Budget:	5902	5902	5865	5186
LRE:	6231	6281	5644	5564

Derived Moderators

Budget Volatility Index: -0.1213 **LRE Volatility Index:** -0.107 **Percent Complete:** 0.0708

BCWS Activity: 0.47671 **BCWP Activity:** 0.57766 **ACWP Activity:** 0.55111

Dependent Variables

Schedule Performance Index: 1.218391 **Cost Performance Index:** 1.70968

Investigator Comments:

Data Identification

Program Tag: A **RatingTag:** A **Project Tag (WBS#):** 3

Project Description: Planning, design, implementation and test of operating system

Rating Information

Rating Date: 10/15/93 **Rating:** 3 **Rating Type:** SPA (EXT) **Rating Relevance:** Med

Rating Comment:

Moderating Variables

Acquisition Phase: Support/Upgrade **Contract Type:** CPI

Program Comments:

S/W Lifecycle: Multiple **Language:** Ada **Language %:** 87.00%

Application: Avionics **Project Budget:** 4201000 **Budget Volatility:** Low

Size: 16300 **% New/Modified Code:** 100.00% **Requirements Volatility:** Unk

Rebaselining : No **Quality Stds On Contract:** ☒ **Quality Params Tracked :** ☒

Cost Accounting Anomalies: Variances may be influenced by letter contract prior to periods of interest

Program Manager Comments:

Cost Data

	Six Months Prior to Rating	Three Months Prior to Rating	Three Months After Rating	Six Months After Rating
Date:	5/30/93	8/30/93	1/30/94	4/30/94
BCWS:	1500	1881	2428	2778
BCWP:	1224	1764	2269	2776
ACWP:	1427	1763	2244	2674
Budget:	5355	5355	5320	4201
LRE:	5140	5143	4820	4066

Derived Moderators

Budget Volatility Index: -0.2155 **LRE Volatility Index:** -0.209 **Percent Complete:** 0.6608

BCWS Activity: 0.46004 **BCWP Activity:** 0.55908 **ACWP Activity:** 0.46634

Dependent Variables

Schedule Performance Index: 1.214397 **Cost Performance Index:** 1.24459

Investigator Comments:

Data Identification

Program Tag: B **RatingTag:** B **Project Tag (WBS#):** 1

Project Description: Analyze, design, and code software for software simulation system component

Rating Information

Rating Date: 1/15/94 **Rating:** 3 **Rating Type:** SPA (EXT) **Rating Relevance:** High

Rating Comment:

Moderating Variables

Acquisition Phase: EMD **Contract Type:** CPAF

Program Comments: May have incentive fee on contract--did not show up in CPR

S/W Lifecycle: Code/Test **Language:** Ada **Language %:** 100.00%

Application: Simulation **Project Budget:** 4300000 **Budget Volatility:** Med

Size: 46746 **% New/Modified Code:** 100.00% **Requirements Volatility:** High

Rebaselining : No **Quality Stds On Contract:** ☒ **Quality Params Tracked :** ☒

Cost Accounting Anomalies: Rebaselining occurrred immediately prior to timeframe of interest. May see repercussions.

Program Manager Comments: The government may be responsible for 50% of the problems ie cost/schedule variances. Contractor has done a "competent job".

Cost Data

	Six Months Prior to Rating	Three Months Prior to Rating	Three Months After Rating	Six Months After Rating
Date:	8/30/93	11/30/93	3/30/94	7/30/94
BCWS:	2807	3016	3524	3823
BCWP:	2794	2985	3349	3763
ACWP:	2829	2959	3353	3765
Budget:	2807	4019	4007	4300
LRE:	2829	4049	4044	4305

Derived Moderators

Budget Volatility Index: 0.53188 **LRE Volatility Index:** 0.5217 **Percent Complete:** 0.8751

BCWS Activity: 0.26576 **BCWP Activity:** 0.25751 **ACWP Activity:** 0.24861

Dependent Variables

Schedule Performance Index: 0.953740 **Cost Performance Index:** 1.03526

Investigator Comments:

Data Identification

Program Tag: B **RatingTag:** B **Project Tag (WBS#):** 2

Project Description: Analyze, design, and code software for software simulation system component

Rating Information

Rating Date: 1/15/94 **Rating:** 3 **Rating Type:** SPA (EXT) **Rating Relevance:** High

Rating Comment:

Moderating Variables

Acquisition Phase: EMD **Contract Type:** CPAF

Program Comments: May have incentive fee on contract--did not show up in CPR

S/W Lifecycle: Code/Test **Language:** Ada **Language %:** 100.00%

Application: Simulation **Project Budget:** 3341000 **Budget Volatility:** Med

Size: 22712 **% New/Modified Code:** 100.00% **Requirements Volatility:** High

Rebaselining : No **Quality Stds On Contract:** ☒ **Quality Params Tracked :** ☒

Cost Accounting Anomalies: Rebaselining occurrred immediately prior to timeframe of interest. May see repercussions.

Program Manager Comments: The government may be responsible for 50% of the problems ie cost/schedule variances. Contractor has done a "competent job".

Cost Data

	Six Months Prior to Rating	Three Months Prior to Rating	Three Months After Rating	Six Months After Rating
Date:	8/30/93	11/30/93	3/30/94	7/30/94
BCWS:	2517	2688	3048	3302
BCWP:	2496	2739	3028	3300
ACWP:	2568	2855	3119	3292
Budget:	2534	3077	3078	3341
LRE:	2597	3146	3214	3365

Derived Moderators

Budget Volatility Index: 0.31847 **LRE Volatility Index:** 0.2957 **Percent Complete:** 0.9877

BCWS Activity: 0.23773 **BCWP Activity:** 0.24364 **ACWP Activity:** 0.21993

Dependent Variables

Schedule Performance Index: 1.024204 **Cost Performance Index:** 1.1105

Investigator Comments:

Data Identification

Program Tag: B   RatingTag: B   Project Tag (WBS#): 3

Project Description: Analyze, design, and code software for software simulation system component

Rating Information

Rating Date: 1/15/94   Rating: 3   Rating Type: SPA (EXT)   Rating Relevance: High

Rating Comment:

Moderating Variables

Acquisition Phase: EMD   Contract Type: CPAF

Program Comments: May have incentive fee on contract--did not show up in CPR

S/W Lifecycle: Code/Test   Language: Ada   Language %: 100.00%

Application: Simulation   Project Budget: 2365000   Budget Volatility: Med

Size: 138837   % New/Modified Code: 100.00%   Requirements Volatility: High

Rebaselining : No   Quality Stds On Contract: ☒   Quality Params Tracked : ☒

Cost Accounting Anomalies: Rebaselining occurrred immediately prior to timeframe of interest. May see repercussions.

Program Manager Comments: The government may be responsible for 50% of the problems ie cost/schedule variances. Contractor has done a "competent job".

Cost Data

	Six Months Prior to Rating	Three Months Prior to Rating	Three Months After Rating	Six Months After Rating
Date:	8/30/93	11/30/93	3/30/94	7/30/94
BCWS:	2039	2151	2194	2315
BCWP:	2036	2129	2183	2314
ACWP:	2075	2176	2276	2337
Budget:	2043	2203	2203	2365
LRE:	2082	2250	2295	2391

Derived Moderators

Budget Volatility Index: 0.15761   LRE Volatility Index: 0.1484   Percent Complete: 0.9784

BCWS Activity: 0.11922   BCWP Activity: 0.12014   ACWP Activity: 0.11211

Dependent Variables

Schedule Performance Index: 1.007246   **Cost Performance Index:** 1.06107

Investigator Comments:

Data Identification

Program Tag: B **RatingTag:** B **Project Tag (WBS#):** 4

Project Description: Analyze, design, and code software for software simulation system component

Rating Information

Rating Date: 1/15/94 **Rating:** 3 **Rating Type:** SPA (EXT) **Rating Relevance:** High

Rating Comment:

Moderating Variables

Acquisition Phase: EMD **Contract Type:** CPAF

Program Comments: May have incentive fee on contract--did not show up in CPR

S/W Lifecycle: Code/Test **Language:** Ada **Language %:** 100.00%

Application: Simulation **Project Budget:** 8685000 **Budget Volatility:** High

Size: 10150 **% New/Modified Code:** 100.00% **Requirements Volatility:** High

Rebaselining : Yes **Quality Stds On Contract:** ☒ **Quality Params Tracked :** ☒

Cost Accounting Anomalies: In Sep 94, a reallocation of budget was detected. Prior to this, they were on budget and on schedule

Program Manager Comments: The government may be responsible for 50% of the problems ie cost/schedule variances. Contractor has done a "competent job".

Cost Data

	Six Months Prior to Rating	Three Months Prior to Rating	Three Months After Rating	Six Months After Rating
Date:	8/30/93	11/30/93	3/30/94	7/30/94
BCWS:	1656	1317	1442	1972
BCWP:	1660	1317	1431	1955
ACWP:	1670	1321	1418	1951
Budget:	1378	-60	2554	8685
LRE:	1393	-46	2573	10085

Derived Moderators

Budget Volatility Index: 5.30261 **LRE Volatility Index:** 6.2398 **Percent Complete:** 0.2251

BCWS Activity: 0.16024 **BCWP Activity:** 0.1509 **ACWP Activity:** 0.14403

Dependent Variables

Schedule Performance Index: 0.933544 **Cost Performance Index:** 1.04982

Investigator Comments:

Data Identification

Program Tag: C **RatingTag:** A **Project Tag (WBS#):** 1

Project Description: Design, code, and test flight control software

Rating Information

Rating Date: 5/15/92 **Rating:** 2 **Rating Type:** SPA (EXT) **Rating Relevance:** High

Rating Comment:

Moderating Variables

Acquisition Phase: Production **Contract Type:** FPIF

Program Comments: 70/30 Share ratio

S/W Lifecycle: Release **Language:** Jovial **Language %:** 100.00%

Application: Avionics **Project Budget:** 3622000 **Budget Volatility:** None

Size: 31000 **% New/Modified Code:** 100.00% **Requirements Volatility:** Low

Rebaselining : No **Quality Stds On Contract:** ☐ **Quality Params Tracked :** ☒

Cost Accounting Anomalies: Minimal effort--Largely complete. May not be enough effort to be a valid data point

Program Manager Comments: Additional requirements & clarifications determined to be in or out of scope. Out-of-scope requirements added as ECPs

Cost Data

	Six Months Prior to Rating	Three Months Prior to Rating	Three Months After Rating	Six Months After Rating
Date:	11/30/91	2/28/92	7/30/92	11/30/92
BCWS:	3532	3532	3538	3538
BCWP:	3539	3539	3545	3545
ACWP:	3705	3716	3716	3716
Budget:	3615	3601	3617	3622
LRE:	3705	3716	3716	3716

Derived Moderators

Budget Volatility Index: 0.00194 **LRE Volatility Index:** 0.003 **Percent Complete:** 0.9787

BCWS Activity: 0.0017 **BCWP Activity:** 0.00169 **ACWP Activity:** 0.00296

Dependent Variables

Schedule Performance Index: 1 **Cost Performance Index:** 0.54545

Investigator Comments:

Datapoint excluded from Complete Data Set due to low activity level.

Data Identification

Program Tag: D **RatingTag:** A **Project Tag (WBS#):** 2

Project Description: Define requirements for each CSCI, perform updates to legacy system

Rating Information

Rating Date: 5/15/91 **Rating:** 1 **Rating Type:** SPA (EXT) **Rating Relevance:** High

Rating Comment: Information provided by Contractor (no program office intermediary)

Moderating Variables

Acquisition Phase: EMD **Contract Type:** CPFF

Program Comments: Program was cancelled.

S/W Lifecycle: Test/Integration **Language:** Jovial **Language %:** 100.00%

Application: Other **Project Budget:** 6282000 **Budget Volatility:** Low

Size: 150000 **% New/Modified Code:** 60.00% **Requirements Volatility:** High

Rebaselining : No **Quality Stds On Contract:** ☒ **Quality Params Tracked :** ☐

Cost Accounting Anomalies:

Program Manager Comments: Program was "overcome by events" and was thus cancelled.

Cost Data

	Six Months Prior to Rating	Three Months Prior to Rating	Three Months After Rating	Six Months After Rating
Date:	12/30/90	3/30/91	8/30/91	11/30/91
BCWS:	3823	4197	4868	5109
BCWP:	3639	4114	4750	4997
ACWP:	4581	5269	5958	6179
Budget:	4445	4850	5000	6282
LRE:	5359	6135	6275	7562

Derived Moderators

Budget Volatility Index: 0.41327 **LRE Volatility Index:** 0.4111 **Percent Complete:** 0.7954

BCWS Activity: 0.25171 **BCWP Activity:** 0.27176 **ACWP Activity:** 0.25862

Dependent Variables

Schedule Performance Index: 1.055988 **Cost Performance Index:** 0.84981

Investigator Comments:

Data Identification

Program Tag: E   RatingTag: A   Project Tag (WBS#): 1

Project Description: Design, code, test, and integration of software for flight control system

Rating Information

Rating Date: 5/15/92   Rating: 2   Rating Type: SPA (EXT)   Rating Relevance: Med

Rating Comment:

Moderating Variables

Acquisition Phase: EMD   Contract Type: CPAF

Program Comments: "Cost plus some base fee plus any incentive (sic) fees awarded"

S/W Lifecycle: Multiple-Early   Language: Ada   Language %: 100.00%

Application: Avionics   Project Budget: 316251000   Budget Volatility: Low

Size: 70000   % New/Modified Code: 100.00%   Requirements Volatility: Low

Rebaselining : No   Quality Stds On Contract: ☒   Quality Params Tracked : ☒

Cost Accounting Anomalies:

Program Manager Comments: Personnel highly experienced in application domain.

Cost Data

	Six Months Prior to Rating	Three Months Prior to Rating	Three Months After Rating	Six Months After Rating
Date:	12/30/91	3/30/92	8/30/92	11/30/92
BCWS:	0	8175	21673	29342
BCWP:	-0	7418	18553	26298
ACWP:	0	7425	19140	28359
Budget:	316251	0	0	316251
LRE:	316251	0	0	316251

Derived Moderators

Budget Volatility Index: 0   LRE Volatility Index: 0   Percent Complete: 0.0832

BCWS Activity: 1   BCWP Activity: 1   ACWP Activity: 1

Dependent Variables

Schedule Performance Index: 0.896258   Cost Performance Index: 0.92732

Investigator Comments:

Data Identification

Program Tag: E **RatingTag:** A **Project Tag (WBS#):** 2

Project Description: Design, code, test, and integration of low-level hardware/software routines for client

Rating Information

Rating Date: 5/15/92 **Rating:** 2 **Rating Type:** SPA (EXT) **Rating Relevance:** Med

Rating Comment:

Moderating Variables

Acquisition Phase: EMD **Contract Type:** CPAF

Program Comments:

S/W Lifecycle: Multiple-Early **Language:** Ada **Language %:** 75.00%

Application: Avionics **Project Budget:** 45545000 **Budget Volatility:** Low

Size: 15000 **% New/Modified Code:** 100.00% **Requirements Volatility:** Med

Rebaselining : No **Quality Stds On Contract:** X **Quality Params Tracked :** X

Cost Accounting Anomalies:

Program Manager Comments: Personnel highly experienced in application domain.

Cost Data

	Six Months Prior to Rating	Three Months Prior to Rating	Three Months After Rating	Six Months After Rating
Date:	12/30/91	3/30/92	8/30/92	11/30/92
BCWS:	0	7238	13302	13756
BCWP:	0	7516	12763	11531
ACWP:	0	6796	11835	12193
Budget:	48634	0	0	45545
LRE:	48634	0	0	45545

Derived Moderators

Budget Volatility Index: -0.0635 **LRE Volatility Index:** -0.064 **Percent Complete:** 0.2532

BCWS Activity: 1 **BCWP Activity:** 1 **ACWP Activity:** 1

Dependent Variables

Schedule Performance Index: 0.838252 **Cost Performance Index:** 0.94571

Investigator Comments:

Data Identification

Program Tag: E **RatingTag:** B **Project Tag (WBS#):** 1

Project Description: Design, code, test, and integration of software for flight control system

Rating Information

Rating Date: 10/15/93 **Rating:** 3 **Rating Type:** SPA (EXT) **Rating Relevance:** High

Rating Comment:

Moderating Variables

Acquisition Phase: EMD **Contract Type:** CPAF

Program Comments:

S/W Lifecycle: Multiple **Language:** Ada **Language %:** 100.00%

Application: Avionics **Project Budget:** 262222000 **Budget Volatility:** High

Size: 70000 **% New/Modified Code:** 100.00% **Requirements Volatility:** Low

Rebaselining : Yes **Quality Stds On Contract:** ☒ **Quality Params Tracked :** ☒

Cost Accounting Anomalies: Rephased during this period

Program Manager Comments: Personnel highly experienced in application domain.

Cost Data

	Six Months Prior to Rating	Three Months Prior to Rating	Three Months After Rating	Six Months After Rating
Date:	5/30/93	8/30/93	1/30/94	4/30/94
BCWS:	46194	54897	69751	79080
BCWP:	43675	52012	68831	77422
ACWP:	43806	51350	64021	73745
Budget:	300751	0	0	262222
LRE:	300751	0	0	250617

Derived Moderators

Budget Volatility Index: -0.1281 **LRE Volatility Index:** -0.167 **Percent Complete:** 0.2953

BCWS Activity: 0.41586 **BCWP Activity:** 0.43588 **ACWP Activity:** 0.40598

Dependent Variables

Schedule Performance Index: 1.026181 **Cost Performance Index:** 1.12719

Investigator Comments:

Note decrease in Budget and LRE during this 12 month period.

Data Identification

Program Tag: E **RatingTag:** B **Project Tag (WBS#):** 2

Project Description: Design, code, test, and integration of low-level hardware/software routines for client

Rating Information

Rating Date: 10/15/93 **Rating:** 3 **Rating Type:** SPA (EXT) **Rating Relevance:** High

Rating Comment:

Moderating Variables

Acquisition Phase: EMD **Contract Type:** CPAF

Program Comments:

S/W Lifecycle: Multiple **Language:** Ada **Language %:** 75.00%

Application: Avionics **Project Budget:** 87704000 **Budget Volatility:** High

Size: 15000 **% New/Modified Code:** 100.00% **Requirements Volatility:** Med

Rebaselining : Yes **Quality Stds On Contract:** ☒ **Quality Params Tracked :** ☒

Cost Accounting Anomalies: Rephased during this period

Program Manager Comments: Personnel highly experienced in application domain.

Cost Data

	Six Months Prior to Rating	Three Months Prior to Rating	Three Months After Rating	Six Months After Rating
Date:	5/30/93	8/30/93	1/30/94	4/30/94
BCWS:	20629	25284	33814	38988
BCWP:	20043	25515	32271	38160
ACWP:	21518	26902	34756	39963
Budget:	61045	0	0	87704
LRE:	61045	0	0	88890

Derived Moderators

Budget Volatility Index: 0.43671 **LRE Volatility Index:** 0.4561 **Percent Complete:** 0.4351

BCWS Activity: 0.47089 **BCWP Activity:** 0.47476 **ACWP Activity:** 0.46155

Dependent Variables

Schedule Performance Index: 0.986818 **Cost Performance Index:** 0.98222

Investigator Comments:

Note decrease in Budget and LRE during this 12 month period.

Data Identification

Program Tag: F **RatingTag:** A **Project Tag (WBS#):** 1

Project Description: Design, develop, code, test, and install 2 Flight Programs, 2 Ground Programs

Rating Information

Rating Date: 11/15/92 **Rating:** 2 **Rating Type:** SPA (EXT) **Rating Relevance:** High

Rating Comment:

Moderating Variables

Acquisition Phase: EMD **Contract Type:** FPIF

Program Comments:

S/W Lifecycle: Code/Test **Language:** Ada **Language %:** 55.60%

Application: Avionics **Project Budget:** 12457000 **Budget Volatility:** Low

Size: 180000 **% New/Modified Code:** 100.00% **Requirements Volatility:** Low

Rebaselining : No **Quality Stds On Contract:** **Quality Params Tracked :**

Cost Accounting Anomalies: Over-target baseline in 1989

Program Manager Comments: Software is in the "top 10" budget drivers, and is a key issue on the program. Subsystems well defined, but there have been integration challenges.

Cost Data

	Six Months Prior to Rating	Three Months Prior to Rating	Three Months After Rating	Six Months After Rating
Date:	5/30/92	8/30/92	1/30/93	4/30/93
BCWS:	11122	11179	11381	11739
BCWP:	10387	10449	10598	10983
ACWP:	16937	17431	17975	18480
Budget:	11231	11231	11381	12457
LRE:	17715	18275	18260	19542

Derived Moderators

Budget Volatility Index: 0.10916 **LRE Volatility Index:** 0.1031 **Percent Complete:** 0.8817

BCWS Activity: 0.05256 **BCWP Activity:** 0.05427 **ACWP Activity:** 0.0835

Dependent Variables

Schedule Performance Index: 0.965964 **Cost Performance Index:** 0.38626

Investigator Comments:

Data Identification

Program Tag: G **RatingTag:** A **Project Tag (WBS#):** 1

Project Description: Software engineering efforts to define, develop, and test system software

Rating Information

Rating Date: 12/15/90 **Rating:** 1 **Rating Type:** SPA (INT) **Rating Relevance:** Very High

Rating Comment:

Moderating Variables

Acquisition Phase: EMD **Contract Type:** FPIF

Program Comments:

S/W Lifecycle: Multiple **Language:** Fortran **Language %:** 61.00%

Application: Command & Co **Project Budget:** 22788000 **Budget Volatility:** Low

Size: 430000 **% New/Modified Code:** 81.00% **Requirements Volatility:** High

Rebaselining : No **Quality Stds On Contract:** ☒ **Quality Params Tracked :** ☒

Cost Accounting Anomalies: Stop work orders, change in direction, etc may affect performance indices

Program Manager Comments: Thinks contractor is a level 2. "Contractor is not as good as some, but better than most"

Cost Data

	Six Months Prior to Rating	Three Months Prior to Rating	Three Months After Rating	Six Months After Rating
Date:	6/30/90	9/30/90	2/28/91	5/30/91
BCWS:	21589	22208	22775	22775
BCWP:	20433	21754	22665	22648
ACWP:	61144	64402	69137	72116
Budget:	22775	22775	22788	22788
LRE:	66623	68767	78000	77549

Derived Moderators

Budget Volatility Index: 0.00057 **LRE Volatility Index:** 0.164 **Percent Complete:** 0.9939

BCWS Activity: 0.05207 **BCWP Activity:** 0.09780 **ACWP Activity:** 0.15214

Dependent Variables

Schedule Performance Index: 1.867622 **Cost Performance Index:** 0.20188

Investigator Comments:

Data Identification

Program Tag: G **RatingTag:** B **Project Tag (WBS#):** 1

Project Description: Software engineering efforts to define, develop, and test system software

Rating Information

Rating Date: 11/15/92 **Rating:** 2 **Rating Type:** SPA (EXT) **Rating Relevance:** Very High

Rating Comment:

Moderating Variables

Acquisition Phase: EMD **Contract Type:** FPIF

Program Comments:

S/W Lifecycle: Multiple **Language:** Fortran **Language %:** 61.00%

Application: Command & Co **Project Budget:** 82378000 **Budget Volatility:** Low

Size: 430000 **% New/Modified Code:** 81.00% **Requirements Volatility:** High

Rebaselining : No **Quality Stds On Contract:** ☒ **Quality Params Tracked :** ☒

Cost Accounting Anomalies: Stop work orders, change in direction, etc may affect performance indices

Program Manager Comments: Thinks contractor is a level 2 "Contractor is not as good as some, better than most"

Cost Data

	Six Months Prior to Rating	Three Months Prior to Rating	Three Months After Rating	Six Months After Rating
Date:	5/30/92	9/30/92	2/28/93	4/30/93
BCWS:	81331	82091	82377	82377
BCWP:	81248	82095	82375	82375
ACWP:	82324	83712	85117	85548
Budget:	81895	82330	82378	82378
LRE:	82692	86042	85431	86463

Derived Moderators

Budget Volatility Index: 0.0059 **LRE Volatility Index:** 0.0456 **Percent Complete:** 1

BCWS Activity: 0.0127 **BCWP Activity:** 0.01368 **ACWP Activity:** 0.03769

Dependent Variables

Schedule Performance Index: 1.077438 **Cost Performance Index:** 0.34957

Investigator Comments:

Data Identification

Program Tag: H   RatingTag: A   Project Tag (WBS#): 1

Project Description: Design, code, test, and integration of software CPCIs

Rating Information

Rating Date: 11/15/92   Rating: 2   Rating Type: SPA (EXT)   Rating Relevance: Very High

Rating Comment:

Moderating Variables

Acquisition Phase: EMD   Contract Type: CPI

Program Comments:

S/W Lifecycle: Multiple   Language: HOL   Language %: 93.00%

Application: Command & Co   Project Budget: 12860000   Budget Volatility: Low

Size: 357714   % New/Modified Code: 69.00%   Requirements Volatility: Med

Rebaselining : No   Quality Stds On Contract: ☒   Quality Params Tracked : ☒

Cost Accounting Anomalies: Internal reallocated effort--"baseline rolling to the right"

Program Manager Comments: Highly concurrent effort. ECPs effectively doubled scope of the effort without stretching schedule--thus increased program schedule risk. Program Manager thinks contractor is level 2. "Not as good as some, but better than most".

Cost Data

	Six Months Prior to Rating	Three Months Prior to Rating	Three Months After Rating	Six Months After Rating
Date:	5/30/92	8/30/92	1/30/93	4/30/93
BCWS:	2863	4294	4879	6178
BCWP:	2652	3736	4879	6124
ACWP:	2334	3681	5251	6483
Budget:	16112	16421	11609	12860
LRE:	16112	16421	11609	12860

Derived Moderators

Budget Volatility Index: -0.2018   LRE Volatility Index: -0.202   Percent Complete: 0.4762

BCWS Activity: 0.53658   BCWP Activity: 0.56695   ACWP Activity: 0.63998

Dependent Variables

Schedule Performance Index: 1.047360   **Cost Performance Index:** 0.83683

Investigator Comments:

Data Identification

Program Tag: I **RatingTag:** A **Project Tag (WBS#):** 1

Project Description: Software-Related management activities: Baselining, Software development planning, etc

Rating Information

Rating Date: 4/15/90 **Rating:** 3 **Rating Type:** SPA (EXT) **Rating Relevance:** High

Rating Comment: SEI conducted the rating

Moderating Variables

Acquisition Phase: Upgrade **Contract Type:** FPIF

Program Comments:

S/W Lifecycle: Multiple-Early **Language:** N/A **Language %:** 0.00%

Application: Database **Project Budget:** 3267000 **Budget Volatility:** Low

Size: 0 **% New/Modified Code:** 0.00% **Requirements Volatility:** Low

Rebaselining : Yes **Quality Stds On Contract:** ☒ **Quality Params Tracked :** ☒

Cost Accounting Anomalies: BCWS decreased in last 6 months of period

Program Manager Comments:

Cost Data

	Six Months Prior to Rating	Three Months Prior to Rating	Three Months After Rating	Six Months After Rating
Date:	10/30/89	1/30/90	6/30/90	9/30/90
BCWS:	2410	2602	2943	2825
BCWP:	2410	2602	2943	2825
ACWP:	2401	2538	2745	2830
Budget:	3083	3083	3083	3267
LRE:	3153	3151	2953	3267

Derived Moderators

Budget Volatility Index: 0.05968 **LRE Volatility Index:** 0.0362 **Percent Complete:** 0.8647

BCWS Activity: 0.14690 **BCWP Activity:** 0.14690 **ACWP Activity:** 0.15159

Dependent Variables

Schedule Performance Index: 1 **Cost Performance Index:** 0.96737

Investigator Comments:

Data Identification

Program Tag: I **RatingTag:** A **Project Tag (WBS#):** 2

Project Description: Specification design and integration oversight tasks. Code and unit test of database architecture.

Rating Information

Rating Date: 4/15/90 **Rating:** 3 **Rating Type:** SPA (EXT) **Rating Relevance:** High

Rating Comment: SEI conducted the rating

Moderating Variables

Acquisition Phase: Upgrade **Contract Type:** FPIF

Program Comments:

S/W Lifecycle: Multiple **Language:** Ada **Language %:** 100.00%

Application: Database **Project Budget:** 4602000 **Budget Volatility:** Low

Size: 40000 **% New/Modified Code:** 15.00% **Requirements Volatility:** Low

Rebaselining : No **Quality Stds On Contract:** ☒ **Quality Params Tracked :** ☒

Cost Accounting Anomalies: Rebaselining prior to this period does not affect this measurement

Program Manager Comments:

Cost Data

	Six Months Prior to Rating	Three Months Prior to Rating	Three Months After Rating	Six Months After Rating
Date:	10/30/89	1/30/90	6/30/90	9/30/90
BCWS:	2244	2306	2568	2621
BCWP:	2178	2226	2452	2620
ACWP:	2169	2340	2582	2724
Budget:	3084	3087	3132	4602
LRE:	3137	3171	3330	4755

Derived Moderators

Budget Volatility Index: 0.49222 **LRE Volatility Index:** 0.5158 **Percent Complete:** 0.5693

BCWS Activity: 0.14384 **BCWP Activity:** 0.16870 **ACWP Activity:** 0.20374

Dependent Variables

Schedule Performance Index: 1.172414 **Cost Performance Index:** 0.7964

Investigator Comments:

Data Identification

Program Tag: I RatingTag: A Project Tag (WBS#): 3

Project Description: Subsystem test, test planning and integration

Rating Information

Rating Date: 4/15/90 Rating: 3 Rating Type: SPA (EXT) Rating Relevance: High

Rating Comment: SEI conducted the rating

Moderating Variables

Acquisition Phase: Upgrade Contract Type: FPIF

Program Comments:

S/W Lifecycle: Test Language: N/A Language %: 0.00%

Application: Database Project Budget: 14880000 Budget Volatility: Low

Size: 0 % New/Modified Code: 0.00% Requirements Volatility: Low

Rebaselining : No Quality Stds On Contract: ☒ Quality Params Tracked : ☒

Cost Accounting Anomalies: Rebaselining prior to this period does not affect this measurement--increase in budget in later qtr.

Program Manager Comments:

Cost Data

	Six Months Prior to Rating	Three Months Prior to Rating	Three Months After Rating	Six Months After Rating
Date:	10/30/89	1/30/90	6/30/90	9/30/90
BCWS:	5189	5892	6881	6948
BCWP:	5038	5652	6739	6949
ACWP:	5029	5698	6635	6968
Budget:	10226	10234	10374	14880
LRE:	11023	10900	11006	15773

Derived Moderators

Budget Volatility Index: 0.45511 LRE Volatility Index: 0.4309 Percent Complete: 0.4670

BCWS Activity: 0.25317 BCWP Activity: 0.27500 ACWP Activity: 0.27827

Dependent Variables

Schedule Performance Index: 1.086413 Cost Performance Index: 0.98556

Investigator Comments:

Data Identification

Program Tag: I **RatingTag:** A **Project Tag (WBS#):** 4

Project Description: Design, code and unit test of CSCIs

Rating Information

Rating Date: 4/15/90 **Rating:** 3 **Rating Type:** SPA (EXT) **Rating Relevance:** High

Rating Comment: SEI conducted the rating

Moderating Variables

Acquisition Phase: Upgrade **Contract Type:** FPIF

Program Comments:

S/W Lifecycle: Multiple **Language:** Ada **Language %:** 100.00%

Application: Database **Project Budget:** 16453000 **Budget Volatility:** Low

Size: 755600 **% New/Modified Code:** 78.00% **Requirements Volatility:** Low

Rebaselining : No **Quality Stds On Contract:** ☒ **Quality Params Tracked :** ☒

Cost Accounting Anomalies: Budget increased in Sept

Program Manager Comments:

Cost Data

	Six Months Prior to Rating	Three Months Prior to Rating	Three Months After Rating	Six Months After Rating
Date:	10/30/89	1/30/90	6/30/90	9/30/90
BCWS:	6253	6671	7638	8821
BCWP:	6013	6496	7210	8817
ACWP:	6379	7245	8675	9300
Budget:	10443	10512	10512	16453
LRE:	11216	11445	12147	17109

Derived Moderators

Budget Volatility Index: 0.57551 **LRE Volatility Index:** 0.5254 **Percent Complete:** 0.5359

BCWS Activity: 0.29112 **BCWP Activity:** 0.31802 **ACWP Activity:** 0.31409

Dependent Variables

Schedule Performance Index: 1.091900 **Cost Performance Index:** 0.95995

Investigator Comments:

Data Identification

Program Tag: I **RatingTag:** A **Project Tag (WBS#):** 5

Project Description: Design, code and unit test of CSCIs

Rating Information

Rating Date: 4/15/90 **Rating:** 3 **Rating Type:** SPA (EXT) **Rating Relevance:** High

Rating Comment: SEI conducted the rating

Moderating Variables

Acquisition Phase: Upgrade **Contract Type:** FPIF

Program Comments:

S/W Lifecycle: Multiple **Language:** Ada **Language %:** 100.00%

Application: Database **Project Budget:** 3822000 **Budget Volatility:** Low

Size: 68000 **% New/Modified Code:** 68.00% **Requirements Volatility:** Low

Rebaselining : No **Quality Stds On Contract:** ☒ **Quality Params Tracked :** ☒

Cost Accounting Anomalies: Budget increased in Sept

Program Manager Comments:

Cost Data

	Six Months Prior to Rating	Three Months Prior to Rating	Three Months After Rating	Six Months After Rating
Date:	10/30/89	1/30/90	6/30/90	9/30/90
BCWS:	2140	2254	2364	2362
BCWP:	2085	2225	2344	2356
ACWP:	2010	2167	2298	2352
Budget:	3056	3077	3077	3822
LRE:	2983	3032	3095	3822

Derived Moderators

Budget Volatility Index: 0.25065 **LRE Volatility Index:** 0.2813 **Percent Complete:** 0.6164

BCWS Activity: 0.09399 **BCWP Activity:** 0.11503 **ACWP Activity:** 0.14541

Dependent Variables

Schedule Performance Index: 1.220721 **Cost Performance Index:** 0.7924

Investigator Comments:

Data Identification

Program Tag: I RatingTag: B Project Tag (WBS#): 1

Project Description: Software-Related management activities: Baselining, Software development planning, etc

Rating Information

Rating Date: 10/15/91 Rating: 1 Rating Type: SCE Rating Relevance: High

Rating Comment:

Moderating Variables

Acquisition Phase: Upgrade Contract Type: Other

Program Comments: contract converted from FPI to FPI/CPFF during this period

S/W Lifecycle: Multiple-Early Language: N/A Language %: 0.00%

Application: Database Project Budget: 2521000 Budget Volatility: Low

Size: 0 % New/Modified Code: 0.00% Requirements Volatility: Med

Rebaselining : No Quality Stds On Contract: Quality Params Tracked :

Cost Accounting Anomalies: Large decrease in budget and actuals. Moved work during this period (Aug 91)--indicated decrease in budget and actuals.

Program Manager Comments:

Cost Data

	Six Months Prior to Rating	Three Months Prior to Rating	Three Months After Rating	Six Months After Rating
Date:	5/30/91	8/30/91	1/30/92	4/30/92
BCWS:	3054	2237	2327	2368
BCWP:	3054	2237	2327	2368
ACWP:	3080	2275	2357	2492
Budget:	3273	2387	2387	2521
LRE:	3334	2438	2429	2693

Derived Moderators

Budget Volatility Index: -0.2298 LRE Volatility Index: -0.192 Percent Complete: 0.9393

BCWS Activity: -0.2897 BCWP Activity: -0.2897 ACWP Activity: -0.236

Dependent Variables

Schedule Performance Index: 1 **Cost Performance Index:** 1.16667

Investigator Comments:

INVALID DATA POINT Accumulated costs (ACWP, BCWP) moved from this project during the period of interest. Invalidates calculation of performance indices.

Data Identification

Program Tag: I **RatingTag:** B **Project Tag (WBS#):** 2

Project Description: Specification design and integration oversight tasks. Code and unit test of database architecture

Rating Information

Rating Date: 10/15/91 **Rating:** 1 **Rating Type:** SCE **Rating Relevance:** High

Rating Comment:

Moderating Variables

Acquisition Phase: Upgrade **Contract Type:** Other

Program Comments: contract converted from FPI to FPI/CPFF during this period

S/W Lifecycle: Multiple **Language:** Ada **Language %:** 100.00%

Application: Database **Project Budget:** 5015000 **Budget Volatility:** Low

Size: 45300 **% New/Modified Code:** 15.00% **Requirements Volatility:** Low

Rebaselining : No **Quality Stds On Contract:** ☒ **Quality Params Tracked :** ☒

Cost Accounting Anomalies:

Program Manager Comments:

Cost Data

	Six Months Prior to Rating	Three Months Prior to Rating	Three Months After Rating	Six Months After Rating
Date:	5/30/91	8/30/91	1/30/92	4/30/92
BCWS:	3390	3724	4157	4466
BCWP:	3296	3646	4099	4322
ACWP:	3471	3840	4405	4679
Budget:	4632	4674	4674	5015
LRE:	4890	5314	5343	5560

Derived Moderators

Budget Volatility Index: 0.08269 **LRE Volatility Index:** 0.1370 **Percent Complete:** 0.8618

BCWS Activity: 0.24093 **BCWP Activity:** 0.23739 **ACWP Activity:** 0.25817

Dependent Variables

Schedule Performance Index: 0.953532 **Cost Performance Index:** 0.84934

Investigator Comments:

Data Identification

Program Tag: I **RatingTag:** B **Project Tag (WBS#):** 3

Project Description: Subsystem test, test planning and integration

Rating Information

Rating Date: 10/15/91 **Rating:** 1 **Rating Type:** SCE **Rating Relevance:** High

Rating Comment:

Moderating Variables

Acquisition Phase: Upgrade **Contract Type:** Other

Program Comments: contract converted from FPI to FPI/CPFF during this period

S/W Lifecycle: Test **Language:** N/A **Language %:** 0.00%

Application: Database **Project Budget:** 15734000 **Budget Volatility:** Low

Size: 0 **% New/Modified Code:** 0.00% **Requirements Volatility:** Low

Rebaselining : No **Quality Stds On Contract:** ☒ **Quality Params Tracked :** ☒

Cost Accounting Anomalies:

Program Manager Comments:

Cost Data

	Six Months Prior to Rating	Three Months Prior to Rating	Three Months After Rating	Six Months After Rating
Date:	5/30/91	8/30/91	1/30/92	4/30/92
BCWS:	8723	9700	11369	12508
BCWP:	8678	9584	11205	12359
ACWP:	8544	9510	11360	12293
Budget:	15008	15122	15219	15734
LRE:	15740	16050	15520	15724

Derived Moderators

Budget Volatility Index: 0.04837 **LRE Volatility Index:** -0.001 **Percent Complete:** 0.7855

BCWS Activity: 0.30261 **BCWP Activity:** 0.29784 **ACWP Activity:** 0.30497

Dependent Variables

Schedule Performance Index: 0.972523 **Cost Performance Index:** 0.98186

Investigator Comments:

Data Identification

Program Tag: I **RatingTag:** B **Project Tag (WBS#):** 4

Project Description: Design, code and unit test of CSCIs

Rating Information

Rating Date: 10/15/91 **Rating:** 1 **Rating Type:** SCE **Rating Relevance:** High

Rating Comment:

Moderating Variables

Acquisition Phase: Upgrade **Contract Type:** Other

Program Comments: contract converted from FPI to FPI/CPFF during this period

S/W Lifecycle: Multiple **Language:** Ada **Language %:** 100.00%

Application: Database **Project Budget:** 17584000 **Budget Volatility:** Low

Size: 874300 **% New/Modified Code:** 78.00% **Requirements Volatility:** Low

Rebaselining : No **Quality Stds On Contract:** ☒ **Quality Params Tracked :** ☒

Cost Accounting Anomalies:

Program Manager Comments:

Cost Data

	Six Months Prior to Rating	Three Months Prior to Rating	Three Months After Rating	Six Months After Rating
Date:	5/30/91	8/30/91	1/30/92	4/30/92
BCWS:	11876	13065	14702	16106
BCWP:	11470	12646	14419	15757
ACWP:	12621	14055	16415	17765
Budget:	16444	16604	16663	17584
LRE:	17657	18632	19066	19889

Derived Moderators

Budget Volatility Index: 0.06933 **LRE Volatility Index:** 0.1264 **Percent Complete:** 0.8961

BCWS Activity: 0.26264 **BCWP Activity:** 0.27207 **ACWP Activity:** 0.28956

Dependent Variables

Schedule Performance Index: 1.013475 **Cost Performance Index:** 0.8334

Investigator Comments:

Data Identification

Program Tag: I | RatingTag: B | Project Tag (WBS#): 5

Project Description: Design, code and unit test of CSCIs

Rating Information

Rating Date: 10/15/91 | Rating: 1 | Rating Type: SCE | Rating Relevance: High

Rating Comment:

Moderating Variables

Acquisition Phase: Upgrade | Contract Type: Other

Program Comments: contract converted from FPI to FPI/CPFF during this period

S/W Lifecycle: Multiple | Language: Ada | Language %: 100.00%

Application: Database | Project Budget: 3953000 | Budget Volatility: Low

Size: 78700 | % New/Modified Code: 68.00% | Requirements Volatility: Low

Rebaselining : No | Quality Stds On Contract: X | Quality Params Tracked : X

Cost Accounting Anomalies:

Program Manager Comments:

Cost Data

	Six Months Prior to Rating	Three Months Prior to Rating	Three Months After Rating	Six Months After Rating
Date:	5/30/91	8/30/91	1/30/92	4/30/92
BCWS:	3041	3354	3722	3848
BCWP:	3017	3261	3595	3810
ACWP:	2973	3223	3664	3946
Budget:	3964	4014	4023	3953
LRE:	3981	4055	4097	4118

Derived Moderators

Budget Volatility Index: -0.0028 | LRE Volatility Index: 0.0344 | Percent Complete: 0.9638

BCWS Activity: 0.20972 | BCWP Activity: 0.20814 | ACWP Activity: 0.24658

Dependent Variables

Schedule Performance Index: 0.982652 | Cost Performance Index: 0.81501

Investigator Comments:

Data Identification

Program Tag: I **RatingTag:** B **Project Tag (WBS#):** 6

Project Description: Software maintenance. Design, code and unit test.

Rating Information

Rating Date: 10/15/91 **Rating:** 1 **Rating Type:** SCE **Rating Relevance:** High

Rating Comment:

Moderating Variables

Acquisition Phase: Upgrade **Contract Type:** Other

Program Comments: contract converted from FPI to FPI/CPFF during this period

S/W Lifecycle: Multiple-Late **Language:** Ada **Language %:** 100.00%

Application: Database **Project Budget:** 1871000 **Budget Volatility:** Low

Size: 0 **% New/Modified Code:** 0.00% **Requirements Volatility:** Low

Rebaselining : No **Quality Stds On Contract:** ☒ **Quality Params Tracked :** ☒

Cost Accounting Anomalies:

Program Manager Comments:

Cost Data

	Six Months Prior to Rating	Three Months Prior to Rating	Three Months After Rating	Six Months After Rating
Date:	5/30/91	8/30/91	1/30/92	4/30/92
BCWS:	0	0	41	203
BCWP:	0	0	23	143
ACWP:	0	0	8	139
Budget:	1074	1074	1074	1871
LRE:	1289	1289	1283	2095

Derived Moderators

Budget Volatility Index: 0.74209 **LRE Volatility Index:** 0.6253 **Percent Complete:** 0.0764

BCWS Activity: 1 **BCWP Activity:** 1 **ACWP Activity:** 1

Dependent Variables

Schedule Performance Index: 0.704433 **Cost Performance Index:** 1.02878

Investigator Comments:

Values for minus 6 month and minus 3 month Budget and LRE are from Oct 91 CPR, which reflects first indication of activity. This was done to avoid DIV 0 errors for derived moderators.

Data Identification

Program Tag: I **RatingTag:** C **Project Tag (WBS#):** 1

Project Description: Software-Related management activities: Baselining, Software development planning, etc

Rating Information

Rating Date: 3/15/93 **Rating:** 1 **Rating Type:** SCE **Rating Relevance:** High

Rating Comment:

Moderating Variables

Acquisition Phase: Upgrade **Contract Type:** Other

Program Comments: contract FPI/CPFF

S/W Lifecycle: Multiple-Early **Language:** N/A **Language %:** 0.00%

Application: **Project Budget:** 2553000 **Budget Volatility:** Low

Size: 0 **% New/Modified Code:** 0.00% **Requirements Volatility:** Med

Rebaselining : No **Quality Stds On Contract:** **Quality Params Tracked :**

Cost Accounting Anomalies: Effort is winding down

Program Manager Comments:

Cost Data

	Six Months Prior to Rating	Three Months Prior to Rating	Three Months After Rating	Six Months After Rating
Date:	9/30/92	12/30/92	5/30/93	8/30/93
BCWS:	2442	2481	2553	2553
BCWP:	2442	2481	2553	2553
ACWP:	2460	2509	2563	2587
Budget:	2558	2558	2553	2553
LRE:	2625	2595	2601	2687

Derived Moderators

Budget Volatility Index: -0.002 **LRE Volatility Index:** 0.0236 **Percent Complete:** 1

BCWS Activity: 0.04348 **BCWP Activity:** 0.04348 **ACWP Activity:** 0.04909

Dependent Variables

Schedule Performance Index: 1 **Cost Performance Index:** 0.87402

Investigator Comments:

Data Identification

Program Tag: I | RatingTag: C | Project Tag (WBS#): 2

Project Description: Specification design and integration oversight tasks. Code and unit test of database architecture

Rating Information

Rating Date: 3/15/93 | Rating: 1 | Rating Type: SCE | Rating Relevance: High

Rating Comment:

Moderating Variables

Acquisition Phase: Upgrade | Contract Type: Other

Program Comments: contract FPI/CPFF

S/W Lifecycle: Multiple | Language: Ada | Language %: 100.00%

Application: Database | Project Budget: 5142000 | Budget Volatility: Low

Size: 54900 | % New/Modified Code: 15.00% | Requirements Volatility: Low

Rebaselining : No | Quality Stds On Contract: ☒ | Quality Params Tracked : ☒

Cost Accounting Anomalies:

Program Manager Comments:

Cost Data

	Six Months Prior to Rating	Three Months Prior to Rating	Three Months After Rating	Six Months After Rating
Date:	9/30/92	12/30/92	5/30/93	8/30/93
BCWS:	4753	4977	5106	5118
BCWP:	4726	4975	5100	5105
ACWP:	5041	5327	5597	5715
Budget:	5156	5156	5142	5142
LRE:	5652	5715	5659	5759

Derived Moderators

Budget Volatility Index: -0.0027 | LRE Volatility Index: 0.0189 | Percent Complete: 0.9928

BCWS Activity: 0.07132 | BCWP Activity: 0.07424 | ACWP Activity: 0.11794

Dependent Variables

Schedule Performance Index: 1.038356 | Cost Performance Index: 0.56231

Investigator Comments:

Data Identification

Program Tag: I **RatingTag:** C **Project Tag (WBS#):** 3

Project Description: Subsystem test, test planning and integration

Rating Information

Rating Date: 3/15/93 **Rating:** 1 **Rating Type:** SCE **Rating Relevance:** High

Rating Comment:

Moderating Variables

Acquisition Phase: Upgrade **Contract Type:** Other

Program Comments: contract FPI/CPFF

S/W Lifecycle: Test **Language:** N/A **Language %:** 0.00%

Application: Database **Project Budget:** 15867000 **Budget Volatility:** Low

Size: 0 **% New/Modified Code:** 0.00% **Requirements Volatility:** Low

Rebaselining : No **Quality Stds On Contract:** X **Quality Params Tracked :** X

Cost Accounting Anomalies:

Program Manager Comments:

Cost Data

	Six Months Prior to Rating	Three Months Prior to Rating	Three Months After Rating	Six Months After Rating
Date:	9/30/92	12/30/92	5/30/93	8/30/93
BCWS:	14279	14761	15363	15730
BCWP:	14204	14654	15274	15668
ACWP:	13708	14388	15126	15455
Budget:	15958	15958	15867	15867
LRE:	15709	15647	15507	15627

Derived Moderators

Budget Volatility Index: -0.0057 **LRE Volatility Index:** -0.005 **Percent Complete:** 0.9875

BCWS Activity: 0.09224 **BCWP Activity:** 0.09344 **ACWP Activity:** 0.11304

Dependent Variables

Schedule Performance Index: 1.008959 **Cost Performance Index:** 0.83801

Investigator Comments:

Data Identification

Program Tag: I **RatingTag:** C **Project Tag (WBS#):** 4

Project Description: Design, code and unit test of CSCIs

Rating Information

Rating Date: 3/15/93 **Rating:** 1 **Rating Type:** SCE **Rating Relevance:** High

Rating Comment:

Moderating Variables

Acquisition Phase: Upgrade **Contract Type:** Other

Program Comments: contract FPI/CPFF

S/W Lifecycle: Multiple **Language:** Ada **Language %:** 100.00%

Application: Database **Project Budget:** 18238000 **Budget Volatility:** Low

Size: 1086000 **% New/Modified Code:** 78.00% **Requirements Volatility:** Low

Rebaselining : No **Quality Stds On Contract:** ☒ **Quality Params Tracked :** ☒

Cost Accounting Anomalies:

Program Manager Comments:

Cost Data

	Six Months Prior to Rating	Three Months Prior to Rating	Three Months After Rating	Six Months After Rating
Date:	9/30/92	12/30/92	5/30/93	8/30/93
BCWS:	17495	17943	18220	18233
BCWP:	17225	17893	18181	18216
ACWP:	19613	20393	21156	21540
Budget:	18286	18263	18238	18238
LRE:	20531	20859	21366	21639

Derived Moderators

Budget Volatility Index: -0.0026 **LRE Volatility Index:** 0.054 **Percent Complete:** 0.9988

BCWS Activity: 0.04048 **BCWP Activity:** 0.05440 **ACWP Activity:** 0.08946

Dependent Variables

Schedule Performance Index: 1.342818 **Cost Performance Index:** 0.51427

Investigator Comments:

Data Identification

Program Tag: I **RatingTag:** C **Project Tag (WBS#):** 5

Project Description: Design, code and unit test of CSCIs

Rating Information

Rating Date: 3/15/93 **Rating:** 1 **Rating Type:** SCE **Rating Relevance:** High

Rating Comment:

Moderating Variables

Acquisition Phase: Upgrade **Contract Type:** Other

Program Comments: contract FPI/CPFF

S/W Lifecycle: Multiple **Language:** Ada **Language %:** 100.00%

Application: Database **Project Budget:** 3951000 **Budget Volatility:** Low

Size: 98000 **% New/Modified Code:** 68.00% **Requirements Volatility:** Low

Rebaselining : No **Quality Stds On Contract:** ☒ **Quality Params Tracked :** ☒

Cost Accounting Anomalies:

Program Manager Comments:

Cost Data

	Six Months Prior to Rating	Three Months Prior to Rating	Three Months After Rating	Six Months After Rating
Date:	9/30/92	12/30/92	5/30/93	8/30/93
BCWS:	3940	3952	3951	3951
BCWP:	3937	3952	3951	3951
ACWP:	4167	4238	4385	4436
Budget:	3951	3952	3951	3951
LRE:	4217	4273	4386	4436

Derived Moderators

Budget Volatility Index: 0 **LRE Volatility Index:** 0.0519 **Percent Complete:** 1

BCWS Activity: 0.00278 **BCWP Activity:** 0.00354 **ACWP Activity:** 0.06064

Dependent Variables

Schedule Performance Index: 1.272727 **Cost Performance Index:** 0.05204

Investigator Comments:

Data point excluded from Complete Data Set due to low activity level.

Data Identification

Program Tag: I **RatingTag:** C **Project Tag (WBS#):** 6

Project Description: Software maintenance. Design, code and unit test.

Rating Information

Rating Date: 3/15/93 **Rating:** 1 **Rating Type:** SCE **Rating Relevance:** High

Rating Comment:

Moderating Variables

Acquisition Phase: Upgrade **Contract Type:** Other

Program Comments: contract FPI/CPFF

S/W Lifecycle: Multiple-Late **Language:** Ada **Language %:** 100.00%

Application: Database **Project Budget:** 2521000 **Budget Volatility:** Low

Size: 0 **% New/Modified Code:** 0.00% **Requirements Volatility:** Low

Rebaselining : No **Quality Stds On Contract:** ☒ **Quality Params Tracked :** ☒

Cost Accounting Anomalies:

Program Manager Comments:

Cost Data

	Six Months Prior to Rating	Three Months Prior to Rating	Three Months After Rating	Six Months After Rating
Date:	9/30/92	12/30/92	5/30/93	8/30/93
BCWS:	1193	1747	2138	2321
BCWP:	1079	1627	2033	2224
ACWP:	904	1334	1870	2076
Budget:	2319	2342	2521	2521
LRE:	2657	2552	2604	2603

Derived Moderators

Budget Volatility Index: 0.08711 **LRE Volatility Index:** -0.020 **Percent Complete:** 0.8822

BCWS Activity: 0.486 **BCWP Activity:** 0.51484 **ACWP Activity:** 0.56455

Dependent Variables

Schedule Performance Index: 1.015071 **Cost Performance Index:** 0.97696

Investigator Comments:

Data Identification

Program Tag: J RatingTag: A Project Tag (WBS#): 1

Project Description: Develop requirements, design, code, and test system software

Rating Information

Rating Date: 3/15/88 Rating: 1 Rating Type: SPA (INT) Rating Relevance: Med

Rating Comment: Government-sponsored contractor did an assessment to suggest possible process improvements

Moderating Variables

Acquisition Phase: EMD Contract Type: FPIF

Program Comments: Similar to previous efforts

S/W Lifecycle: Requirements Language: Jovial Language %: 100.00%

Application: Command & Co Project Budget: 7488000 Budget Volatility: Low

Size: 148000 % New/Modified Code: 100.00% Requirements Volatility: Low

Rebaselining : No Quality Stds On Contract: ☐ Quality Params Tracked : ☒

Cost Accounting Anomalies: None

Program Manager Comments: Beat target sched. Had experience with previous similar project, but subcontracted the software development. Fell behind early in project, but instituted process improvement initiatives and got well. Size in DSI

Cost Data

	Six Months Prior to Rating	Three Months Prior to Rating	Three Months After Rating	Six Months After Rating
Date:	4/30/88	7/30/88	12/30/88	3/30/89
BCWS:	0	0	100	675
BCWP:	0	0	86	493
ACWP:	0	0	80	486
Budget:	7488	7488	7488	7488
LRE:	7488	7488	7488	7492

Derived Moderators

Budget Volatility Index: 0 LRE Volatility Index: 0.0005 Percent Complete: 0.0658

BCWS Activity: 1 BCWP Activity: 1 ACWP Activity: 1

Dependent Variables

Schedule Performance Index: 0.730370 **Cost Performance Index:** 1.01440

Investigator Comments:

Values for minus 6 month and minus 3 month Budget and LRE are from Dec 88 CPR. This was done to avoid DIV 0 errors for derived moderators. Program initiated at time organization was rated. Data representative of 12 months after

Data Identification

Program Tag: J **RatingTag:** A **Project Tag (WBS#):** 2

Project Description: Develop requirements, design, code, and test system software

Rating Information

Rating Date: 3/15/88 **Rating:** 1 **Rating Type:** SPA (INT) **Rating Relevance:** Med

Rating Comment: Government-sponsored contractor did an assessment to suggest possible process improvements

Moderating Variables

Acquisition Phase: EMD **Contract Type:** FPIF

Program Comments: Similar to previous efforts

S/W Lifecycle: Test/Integration **Language:** Jovial **Language %:** 100.00%

Application: Simulation **Project Budget:** 2557000 **Budget Volatility:** Low

Size: 42000 **% New/Modified Code:** 52.00% **Requirements Volatility:** Low

Rebaselining : No **Quality Stds On Contract:** ☐ **Quality Params Tracked :** ☒

Cost Accounting Anomalies:

Program Manager Comments: Beat target sched. Had experience with previous similar project, but subcontracted the software development. Fell behind early in project, but instituted process improvement initiatives and got well. Size in DSI

Cost Data

	Six Months Prior to Rating	Three Months Prior to Rating	Three Months After Rating	Six Months After Rating
Date:	4/30/88	7/30/88	12/30/88	3/30/89
BCWS:	0	0	89	360
BCWP:	0	0	19	109
ACWP:	0	0	20	107
Budget:	2557	2557	2557	2557
LRE:	2557	2557	2557	2557

Derived Moderators

Budget Volatility Index: 0 **LRE Volatility Index:** 0 **Percent Complete:** 0.0426

BCWS Activity: 1 **BCWP Activity:** 1 **ACWP Activity:** 1

Dependent Variables

Schedule Performance Index: 0.302778 **Cost Performance Index:** 1.01869

Investigator Comments:

Values for minus 6 month and minus 3 month Budget and LRE are from Dec 88 CPR. This was done to avoid DIV 0 errors for derived moderators. Program initiated at time organization was rated. Data representative of 12 months after rating

Data Identification

Program Tag: J **RatingTag:** A **Project Tag (WBS#):** 3

Project Description: Develop requirements, design, code, and test system software

Rating Information

Rating Date: 3/15/88 **Rating:** 1 **Rating Type:** SPA (INT) **Rating Relevance:** Med

Rating Comment: Government-sponsored contractor did an assessment to suggest possible process improvements

Moderating Variables

Acquisition Phase: EMD **Contract Type:** FPIF

Program Comments: Similar to previous efforts

S/W Lifecycle: Requirements **Language:** Fortran **Language %:** 100.00%

Application: Command & Co **Project Budget:** 3283000 **Budget Volatility:** Low

Size: 141000 **% New/Modified Code:** 91.00% **Requirements Volatility:** Low

Rebaselining : No **Quality Stds On Contract:** ☐ **Quality Params Tracked :** ☒

Cost Accounting Anomalies: Subcontracting plan did not materialize--thus more effort expended than budgeted

Program Manager Comments: Beat target sched. Had experience with previous similar project, but subcontracted the software development. Fell behind early in project, but instituted process improvement initiatives and got well. Size in DSI

Cost Data

	Six Months Prior to Rating	Three Months Prior to Rating	Three Months After Rating	Six Months After Rating
Date:	4/30/88	7/30/88	12/30/88	3/30/89
BCWS:	0	0	189	518
BCWP:	0	0	161	452
ACWP:	0	0	164	419
Budget:	3284	3284	3284	3284
LRE:	3284	3284	3284	3283

Derived Moderators

Budget Volatility Index: 0 **LRE Volatility Index:** -0.0003 **Percent Complete:** 0.1376

BCWS Activity: 1 **BCWP Activity:** 1 **ACWP Activity:** 1

Dependent Variables

Schedule Performance Index: 0.872587 **Cost Performance Index:** 1.07876

Investigator Comments:

Values for minus 6 month and minus 3 month Budget and LRE are from Dec 88 CPR. This was done to avoid DIV 0 errors for derived moderators. Program initiated at time organization was rated. Data representative of 12 months after rating

Data Identification

Program Tag: J **RatingTag:** B **Project Tag (WBS#):** 1

Project Description: Develop requirements, design, code, and test system software

Rating Information

Rating Date: 4/15/91 **Rating:** 3 **Rating Type:** SCE **Rating Relevance:** Med

Rating Comment:

Moderating Variables

Acquisition Phase: EMD **Contract Type:** FPIF

Program Comments: Similar to previous efforts

S/W Lifecycle: Test/Integration **Language:** Jovial **Language %:** 100.00%

Application: Command & Co **Project Budget:** 7998000 **Budget Volatility:** Low

Size: 148000 **% New/Modified Code:** 100.00% **Requirements Volatility:** Low

Rebaselining : No **Quality Stds On Contract:** ☐ **Quality Params Tracked :** ☒

Cost Accounting Anomalies:

Program Manager Comments: Beat target sched. Size in DSI

Cost Data

	Six Months Prior to Rating	Three Months Prior to Rating	Three Months After Rating	Six Months After Rating
Date:	10/30/90	1/30/91	6/30/91	9/30/91
BCWS:	6521	7255	7928	7998
BCWP:	6671	7260	7853	8000
ACWP:	6962	7697	8198	8207
Budget:	7930	7985	7998	7998
LRE:	7820	7985	8201	8201

Derived Moderators

Budget Volatility Index: 0.00858 **LRE Volatility Index:** 0.0487 **Percent Complete:** 1.0003

BCWS Activity: 0.18467 **BCWP Activity:** 0.16613 **ACWP Activity:** 0.1517

Dependent Variables

Schedule Performance Index: 0.899797 **Cost Performance Index:** 1.06747

Investigator Comments:

Data Identification

Program Tag: J   RatingTag: B   Project Tag (WBS#): 2

Project Description: Develop requirements, design, code, and test system software

Rating Information

Rating Date: 4/15/91   Rating: 3   Rating Type: SCE   Rating Relevance: Med

Rating Comment:

Moderating Variables

Acquisition Phase: EMD   Contract Type: FPIF

Program Comments: Similar to previous efforts

S/W Lifecycle: Test/Integration   Language: Jovial   Language %: 100.00%

Application: Simulation   Project Budget: 2654000   Budget Volatility: Low

Size: 42000   % New/Modified Code: 52.00%   Requirements Volatility: Low

Rebaselining : No   Quality Stds On Contract: ☐   Quality Params Tracked : ☒

Cost Accounting Anomalies:

Program Manager Comments: Beat target sched. Size in DSI

Cost Data

	Six Months Prior to Rating	Three Months Prior to Rating	Three Months After Rating	Six Months After Rating
Date:	10/30/90	1/30/91	6/30/91	9/30/91
BCWS:	2315	2450	2628	2654
BCWP:	2217	2382	2628	2655
ACWP:	2015	2152	2235	2236
Budget:	2654	2654	2654	2654
LRE:	2563	2320	2235	2235

Derived Moderators

Budget Volatility Index: 0   LRE Volatility Index: -0.128   Percent Complete: 1.0004

BCWS Activity: 0.12773   BCWP Activity: 0.16497   ACWP Activity: 0.09884

Dependent Variables

Schedule Performance Index: 1.292035   **Cost Performance Index:** 1.98190

Investigator Comments:

Data Identification

Program Tag: J **RatingTag:** B **Project Tag (WBS#):** 3

Project Description: Develop requirements, design, code, and test system software

Rating Information

Rating Date: 4/30/91 **Rating:** 3 **Rating Type:** SCE **Rating Relevance:** Med

Rating Comment:

Moderating Variables

Acquisition Phase: EMD **Contract Type:** FPIF

Program Comments: Similar to previous efforts

S/W Lifecycle: Test/Integration **Language:** Fortran **Language %:** 100.00%

Application: Command & Co **Project Budget:** 3432000 **Budget Volatility:** Low

Size: 141000 **% New/Modified Code:** 91.00% **Requirements Volatility:** Low

Rebaselining : No **Quality Stds On Contract:** ☐ **Quality Params Tracked :** ☒

Cost Accounting Anomalies:

Program Manager Comments: Beat target sched. Size in DSI

Cost Data

	Six Months Prior to Rating	Three Months Prior to Rating	Three Months After Rating	Six Months After Rating
Date:	10/30/90	1/30/91	6/30/91	9/30/91
BCWS:	3009	3171	3400	3432
BCWP:	2880	3086	3406	3433
ACWP:	3252	3395	3506	3508
Budget:	3432	3432	3432	3432
LRE:	3497	3497	3513	3507

Derived Moderators

Budget Volatility Index: 0 **LRE Volatility Index:** 0.0029 **Percent Complete:** 1.0003

BCWS Activity: 0.12325 **BCWP Activity:** 0.16108 **ACWP Activity:** 0.07298

Dependent Variables

Schedule Performance Index: 1.307329 **Cost Performance Index:** 2.16016

Investigator Comments:

Data Identification

Program Tag: J RatingTag: C Project Tag (WBS#): 1

Project Description: Develop requirements, design, code, and test system software

Rating Information

Rating Date: 11/15/91 Rating: 3 Rating Type: SCE Rating Relevance: High

Rating Comment:

Moderating Variables

Acquisition Phase: EMD Contract Type: FPIF

Program Comments: Similar to previous efforts

S/W Lifecycle: Integration Language: Jovial Language %: 100.00%

Application: Command & Co Project Budget: 7998000 Budget Volatility: Low

Size: 148000 % New/Modified Code: 100.00% Requirements Volatility: Low

Rebaselining : No Quality Stds On Contract: ☐ Quality Params Tracked : ☒

Cost Accounting Anomalies: Very little effort over the period of interest--Actuals over period only .3% of actuals to date--will affect CPI

Program Manager Comments: Beat target sched. Size in DSI

Cost Data

	Six Months Prior to Rating	Three Months Prior to Rating	Three Months After Rating	Six Months After Rating
Date:	5/30/91	8/30/91	1/30/92	4/30/92
BCWS:	7852	7998	7998	7998
BCWP:	7769	7998	7997	7998
ACWP:	8171	8201	8195	8195
Budget:	7998	7998	7998	7998
LRE:	8186	8201	8204	8195

Derived Moderators

Budget Volatility Index: 0 LRE Volatility Index: 0.0011 Percent Complete: 1

BCWS Activity: 0.01825 BCWP Activity: 0.02863 ACWP Activity: 0.00293

Dependent Variables

Schedule Performance Index: 1.568493 **Cost Performance Index:** 9.54167

Investigator Comments:

Data point excluded from Complete Data Set due to low activity level.

Data Identification

Program Tag: J   RatingTag: C   Project Tag (WBS#): 2

Project Description: Develop requirements, design, code, and test system software

Rating Information

Rating Date: 11/15/91   Rating: 3   Rating Type: SCE   Rating Relevance: High

Rating Comment:

Moderating Variables

Acquisition Phase: EMD   Contract Type: FPIF

Program Comments: Similar to previous efforts

S/W Lifecycle: Integration   Language: Jovial   Language %: 100.00%

Application: Simulation   Project Budget: 2654000   Budget Volatility:

Size: 0   % New/Modified Code: 52.00%   Requirements Volatility: Low

Rebaselining : No   Quality Stds On Contract: ☐   Quality Params Tracked : ☒

Cost Accounting Anomalies: No effort for this WBS over the time period of interest--may affect performance indices

Program Manager Comments: Beat target sched. Size in DSI

Cost Data

	Six Months Prior to Rating	Three Months Prior to Rating	Three Months After Rating	Six Months After Rating
Date:	5/30/91	8/30/91	1/30/92	4/30/92
BCWS:	2605	2654	2654	2654
BCWP:	2605	2654	2654	2654
ACWP:	2233	2235	2233	2233
Budget:	2654	2654	2654	2654
LRE:	2235	2235	2235	2233

Derived Moderators

Budget Volatility Index: 0   LRE Volatility Index: -0.0009   Percent Complete: 1

BCWS Activity: 0.01846   BCWP Activity: 0.01846   ACWP Activity: 0

Dependent Variables

Schedule Performance Index: 1   Cost Performance Index: #Error

Investigator Comments:

Data point excluded from Complete Data Set due to low activity level.

Data Identification

Program Tag: J **RatingTag:** C **Project Tag (WBS#):** 3

Project Description: Develop requirements, design, code, and test system software

Rating Information

Rating Date: 11/15/91 **Rating:** 3 **Rating Type:** SCE **Rating Relevance:** High

Rating Comment:

Moderating Variables

Acquisition Phase: EMD **Contract Type:** FPIF

Program Comments: Similar to previous efforts

S/W Lifecycle: Integration **Language:** Fortran **Language %:** 100.00%

Application: Command & Co **Project Budget:** 3432000 **Budget Volatility:** Low

Size: 141000 **% New/Modified Code:** 91.00% **Requirements Volatility:** Low

Rebaselining : No **Quality Stds On Contract:** ☐ **Quality Params Tracked :** ☒

Cost Accounting Anomalies: Little effort for this WBS over the time period of interest--may affect performance indices

Program Manager Comments: Beat target sched. Size in DSI

Cost Data

	Six Months Prior to Rating	Three Months Prior to Rating	Three Months After Rating	Six Months After Rating
Date:	5/30/91	8/30/91	1/30/92	4/30/92
BCWS:	3366	3432	3432	3432
BCWP:	3363	3432	3431	3432
ACWP:	3493	3507	3506	3506
Budget:	3432	3432	3432	3432
LRE:	3513	3507	3507	3506

Derived Moderators

Budget Volatility Index: 0 **LRE Volatility Index:** -0.002 **Percent Complete:** 1

BCWS Activity: 0.01923 **BCWP Activity:** 0.02010 **ACWP Activity:** 0.00371

Dependent Variables

Schedule Performance Index: 1.045455 **Cost Performance Index:** 5.30769

Investigator Comments:

Data point excluded from Complete Data Set due to low activity level.

Data Identification

Program Tag: K **RatingTag:** A **Project Tag (WBS#):** 2

Project Description: Subsystem architecture, database administration, and software configuration management.

Rating Information

Rating Date: 12/15/89 **Rating:** 2 **Rating Type:** SPA (INT) **Rating Relevance:** High

Rating Comment:

Moderating Variables

Acquisition Phase: Support/Upgrade **Contract Type:** FPIF

Program Comments:

S/W Lifecycle: Multiple **Language:** N/A **Language %:** 0.00%

Application: Database **Project Budget:** 8451000 **Budget Volatility:** Low

Size: 0 **% New/Modified Code:** 0.00% **Requirements Volatility:** Low

Rebaselining : No **Quality Stds On Contract:** ☐ **Quality Params Tracked :** ☒

Cost Accounting Anomalies: No +/- three month data

Program Manager Comments:

Cost Data

	Six Months Prior to Rating	Three Months Prior to Rating	Three Months After Rating	Six Months After Rating
Date:	6/30/89			5/30/90
BCWS:	6767	0		7863
BCWP:	6755	0		7821
ACWP:	7060	0		8288
Budget:	7475	0		8451
LRE:	7684	0		8714

Derived Moderators

Budget Volatility Index: 0.13057 **LRE Volatility Index:** 0.1340 **Percent Complete:** 0.9255

BCWS Activity: 0.13939 **BCWP Activity:** 0.1363 **ACWP Activity:** 0.14817

Dependent Variables

Schedule Performance Index: 0.972628 **Cost Performance Index:** 0.86808

Investigator Comments:

No data for plus/minus three month.

Data Identification

Program Tag: K **RatingTag:** A **Project Tag (WBS#):** 3

Project Description: Overall mangement of software development effort

Rating Information

Rating Date: 12/15/89 **Rating:** 2 **Rating Type:** SPA (INT) **Rating Relevance:** High

Rating Comment:

Moderating Variables

Acquisition Phase: Support/Upgrade **Contract Type:** FPIF

Program Comments:

S/W Lifecycle: Multiple **Language:** N/A **Language %:** 0.00%

Application: Database **Project Budget:** 3205000 **Budget Volatility:** Low

Size: 0 **% New/Modified Code:** 0.00% **Requirements Volatility:** Low

Rebaselining : No **Quality Stds On Contract:** ☐ **Quality Params Tracked :** ☒

Cost Accounting Anomalies: No +/- three month data

Program Manager Comments:

Cost Data

	Six Months Prior to Rating	Three Months Prior to Rating	Three Months After Rating	Six Months After Rating
Date:	6/30/89			5/30/90
BCWS:	2025	0	0	2824
BCWP:	2025	0	0	2824
ACWP:	2071	0	0	2727
Budget:	2237	0	0	3205
LRE:	2334	0	0	3351

Derived Moderators

Budget Volatility Index: 0.43272 **LRE Volatility Index:** 0.4357 **Percent Complete:** 0.8811

BCWS Activity: 0.28293 **BCWP Activity:** 0.28293 **ACWP Activity:** 0.24056

Dependent Variables

Schedule Performance Index: 1 **Cost Performance Index:** 1.21799

Investigator Comments:

No data for plus/minus three month.

Data Identification

Program Tag: K RatingTag: A Project Tag (WBS#): 4

Project Description: Requirements, design, code, and test of system control CSCI

Rating Information

Rating Date: 12/15/89 Rating: 2 Rating Type: SPA (INT) Rating Relevance: High

Rating Comment:

Moderating Variables

Acquisition Phase: Support/Upgrade Contract Type: FPIF

Program Comments:

S/W Lifecycle: Multiple Language: Fortran Language %: 100.00%

Application: Database Project Budget: 2440000 Budget Volatility: Low

Size: 22400 % New/Modified Code: 85.00% Requirements Volatility: Low

Rebaselining : No Quality Stds On Contract: ☐ Quality Params Tracked : ☒

Cost Accounting Anomalies: No +/- three month data

Program Manager Comments:

Cost Data

	Six Months Prior to Rating	Three Months Prior to Rating	Three Months After Rating	Six Months After Rating
Date:	6/30/89			5/30/89
BCWS:	2158	0	0	2440
BCWP:	2160	0	0	2416
ACWP:	2158	0	0	2334
Budget:	2415	0	0	2440
LRE:	2412	0	0	2437

Derived Moderators

Budget Volatility Index: 0.01035 LRE Volatility Index: 0.0104 Percent Complete: 0.9902

BCWS Activity: 0.11557 BCWP Activity: 0.10596 ACWP Activity: 0.07541

Dependent Variables

Schedule Performance Index: 0.907801 Cost Performance Index: 1.45455

Investigator Comments:

No data for plus/minus three month.

Data Identification

Program Tag: K | RatingTag: A | Project Tag (WBS#): 5

Project Description: Requirements, design, code, and test of systems interface CSCI

Rating Information

Rating Date: 12/15/89 | Rating: 2 | Rating Type: SPA (INT) | Rating Relevance: High

Rating Comment:

Moderating Variables

Acquisition Phase: Support/Upgrade | Contract Type: FPIF

Program Comments:

S/W Lifecycle: Multiple | Language: Fortran | Language %: 100.00%

Application: Database | Project Budget: 4238000 | Budget Volatility: Low

Size: 43200 | % New/Modified Code: 85.00% | Requirements Volatility: Low

Rebaselining : No | Quality Stds On Contract: ☐ | Quality Params Tracked : ☒

Cost Accounting Anomalies: No +/- three month data

Program Manager Comments:

Cost Data

	Six Months Prior to Rating	Three Months Prior to Rating	Three Months After Rating	Six Months After Rating
Date:	6/30/89			5/30/90
BCWS:	2286	0	0	3268
BCWP:	2279	0	0	3167
ACWP:	2190	0	0	2989
Budget:	2581	0	0	4238
LRE:	2515	0	0	4169

Derived Moderators

Budget Volatility Index: 0.642 | LRE Volatility Index: 0.6577 | Percent Complete: 0.7473

BCWS Activity: 0.30049 | BCWP Activity: 0.28039 | ACWP Activity: 0.26731

Dependent Variables

Schedule Performance Index: 0.904277 | **Cost Performance Index:** 1.11139

Investigator Comments:

No data for plus/minus three month.

Data Identification

Program Tag: K RatingTag: A Project Tag (WBS#): 6

Project Description: Requirements, design, code, and test of applications CSCI

Rating Information

Rating Date: 12/15/89 Rating: 2 Rating Type: SPA (INT) Rating Relevance: High

Rating Comment:

Moderating Variables

Acquisition Phase: Support/Upgrade Contract Type: FPIF

Program Comments:

S/W Lifecycle: Multiple Language: Fortran Language %: 100.00%

Application: Database Project Budget: 2683000 Budget Volatility: Low

Size: 73200 % New/Modified Code: 85.00% Requirements Volatility: Low

Rebaselining : No Quality Stds On Contract: ☐ Quality Params Tracked : ☒

Cost Accounting Anomalies: No +/- three month data

Program Manager Comments:

Cost Data

	Six Months Prior to Rating	Three Months Prior to Rating	Three Months After Rating	Six Months After Rating
Date:	6/30/89			5/30/90
BCWS:	2424	0	0	2683
BCWP:	2418	0	0	2655
ACWP:	2510	0	0	2645
Budget:	2516	0	0	2683
LRE:	2609	0	0	2755

Derived Moderators

Budget Volatility Index: 0.06638 LRE Volatility Index: 0.056 Percent Complete: 0.9896

BCWS Activity: 0.09653 BCWP Activity: 0.08927 ACWP Activity: 0.05104

Dependent Variables

Schedule Performance Index: 0.915058 Cost Performance Index: 1.75556

Investigator Comments:

No data for plus/minus three month.

Data Identification

Program Tag: K **RatingTag:** A **Project Tag (WBS#):** 7

Project Description: Requirements, design, code, and test of database maintenance CSCI

Rating Information

Rating Date: 12/15/89 **Rating:** 2 **Rating Type:** SPA (INT) **Rating Relevance:** High

Rating Comment:

Moderating Variables

Acquisition Phase: Support/Upgrade **Contract Type:** FPIF

Program Comments:

S/W Lifecycle: Multiple **Language:** Fortran **Language %:** 100.00%

Application: Database **Project Budget:** 2667000 **Budget Volatility:** Low

Size: 25700 **% New/Modified Code:** 85.00% **Requirements Volatility:** Low

Rebaselining : No **Quality Stds On Contract:** ☐ **Quality Params Tracked :** ☒

Cost Accounting Anomalies: No +/- three month data

Program Manager Comments:

Cost Data

	Six Months Prior to Rating	Three Months Prior to Rating	Three Months After Rating	Six Months After Rating
Date:	6/30/89			5/30/90
BCWS:	2486	0	0	2650
BCWP:	2488	0	0	2650
ACWP:	2787	0	0	2866
Budget:	2616	0	0	2667
LRE:	2991	0	0	2874

Derived Moderators

Budget Volatility Index: 0.0195 **LRE Volatility Index:** -0.039 **Percent Complete:** 0.9936

BCWS Activity: 0.06189 **BCWP Activity:** 0.06113 **ACWP Activity:** 0.02756

Dependent Variables

Schedule Performance Index: 0.987805 **Cost Performance Index:** 2.05063

Investigator Comments:

No data for plus/minus three month.

Data Identification

Program Tag: K **RatingTag:** A **Project Tag (WBS#):** 8

Project Description: Requirements, design, code, and test of database support CSCI

Rating Information

Rating Date: 12/15/89 **Rating:** 2 **Rating Type:** SPA (INT) **Rating Relevance:** High

Rating Comment:

Moderating Variables

Acquisition Phase: Support/Upgrade **Contract Type:** FPIF

Program Comments:

S/W Lifecycle: Multiple **Language:** Fortran **Language %:** 100.00%

Application: Database **Project Budget:** 1181000 **Budget Volatility:** Low

Size: 14200 **% New/Modified Code:** 85.00% **Requirements Volatility:** Low

Rebaselining : No **Quality Stds On Contract:** ☐ **Quality Params Tracked :** ☒

Cost Accounting Anomalies: No +/- three month data

Program Manager Comments:

Cost Data

	Six Months Prior to Rating	Three Months Prior to Rating	Three Months After Rating	Six Months After Rating
Date:	6/30/89			5/30/90
BCWS:	1162	0	0	1175
BCWP:	1160	0	0	1175
ACWP:	1258	0	0	1266
Budget:	1162	0	0	1181
LRE:	1262	0	0	1277

Derived Moderators

Budget Volatility Index: 0.01635 **LRE Volatility Index:** 0.0119 **Percent Complete:** 0.9949

BCWS Activity: 0.01106 **BCWP Activity:** 0.01277 **ACWP Activity:** 0.00632

Dependent Variables

Schedule Performance Index: 1.153846 **Cost Performance Index:** 1.875

Investigator Comments:

Data point excluded from Complete Data Set due to low activity level. No data for plus/minus three months.

Data Identification

Program Tag: K **RatingTag:** A **Project Tag (WBS#):** 9

Project Description: Software integration activities.

Rating Information

Rating Date: 12/15/89 **Rating:** 2 **Rating Type:** SPA (INT) **Rating Relevance:** High

Rating Comment:

Moderating Variables

Acquisition Phase: Support/Upgrade **Contract Type:** FPIF

Program Comments:

S/W Lifecycle: Test/Integration **Language:** Fortran **Language %:** 100.00%

Application: Database **Project Budget:** 5821000 **Budget Volatility:** Low

Size: 0 **% New/Modified Code:** 0.00% **Requirements Volatility:** Low

Rebaselining : No **Quality Stds On Contract:** ☐ **Quality Params Tracked :** ☒

Cost Accounting Anomalies: No +/- three month data

Program Manager Comments:

Cost Data

	Six Months Prior to Rating	Three Months Prior to Rating	Three Months After Rating	Six Months After Rating
Date:	6/30/89			5/30/90
BCWS:	3009	0	0	4949
BCWP:	3002	0	0	4784
ACWP:	5287	0	0	7574
Budget:	5928	0	0	5821
LRE:	7906	0	0	8375

Derived Moderators

Budget Volatility Index: -0.0180 **LRE Volatility Index:** 0.0593 **Percent Complete:** 0.8219

BCWS Activity: 0.392 **BCWP Activity:** 0.37249 **ACWP Activity:** 0.30195

Dependent Variables

Schedule Performance Index: 0.918557 **Cost Performance Index:** 0.77919

Investigator Comments:

No data for plus/minus three month.

Data Identification

Program Tag: K RatingTag: B Project Tag (WBS#): 2

Project Description: Subsystem architecture, database administration, and software configuration management.

Rating Information

Rating Date: 9/15/90 Rating: 2 Rating Type: SCE Rating Relevance: High

Rating Comment:

Moderating Variables

Acquisition Phase: Contract Type:

Program Comments:

S/W Lifecycle: Multiple Language: N/A Language %: 0.00%

Application: Database Project Budget: 8586000 Budget Volatility: Low

Size: 0 % New/Modified Code: 0.00% Requirements Volatility: Low

Rebaselining : No Quality Stds On Contract: ☐ Quality Params Tracked : ☒

Cost Accounting Anomalies: No +/- three month data

Program Manager Comments:

Cost Data

	Six Months Prior to Rating	Three Months Prior to Rating	Three Months After Rating	Six Months After Rating
Date:	3/30/90	6/30/90	11/30/90	2/28/91
BCWS:	7675	0		8503
BCWP:	7647	0		8490
ACWP:	8078	0		9002
Budget:	8451	0		8586
LRE:	8695	0		9122

Derived Moderators

Budget Volatility Index: 0.01597 LRE Volatility Index: 0.0491 Percent Complete: 0.9888

BCWS Activity: 0.09738 BCWP Activity: 0.09929 ACWP Activity: 0.10264

Dependent Variables

Schedule Performance Index: 1.018116 Cost Performance Index: 0.91234

Investigator Comments:

No data for plus/minus three month.

Data Identification

Program Tag: K RatingTag: B Project Tag (WBS#): 3

Project Description: Overall mangement of software development effort

Rating Information

Rating Date: 9/15/90 Rating: 2 Rating Type: SCE Rating Relevance: High

Rating Comment:

Moderating Variables

Acquisition Phase: Contract Type:

Program Comments:

S/W Lifecycle: Multiple Language: N/A Language %: 0.00%

Application: Database Project Budget: 3239000 Budget Volatility: Low

Size: 0 % New/Modified Code: 0.00% Requirements Volatility: Low

Rebaselining : No Quality Stds On Contract: ☐ Quality Params Tracked : ☒

Cost Accounting Anomalies: No +/- three month data

Program Manager Comments:

Cost Data

	Six Months Prior to Rating	Three Months Prior to Rating	Three Months After Rating	Six Months After Rating
Date:	3/30/90	6/30/90	11/30/90	2/28/91
BCWS:	2679	0	0	3211
BCWP:	2679	0	0	3211
ACWP:	2609	0	0	3116
Budget:	3205	0	0	3239
LRE:	3351	0	0	3197

Derived Moderators

Budget Volatility Index: 0.01061 LRE Volatility Index: -0.046 Percent Complete: 0.9914

BCWS Activity: 0.16568 BCWP Activity: 0.16568 ACWP Activity: 0.16271

Dependent Variables

Schedule Performance Index: 1 Cost Performance Index: 1.04931

Investigator Comments:

No data for plus/minus three month.

Data Identification

Program Tag: K **RatingTag:** B **Project Tag (WBS#):** 4

Project Description: Requirements, design, code, and test of system control CSCI

Rating Information

Rating Date: 9/15/90 **Rating:** 2 **Rating Type:** SCE **Rating Relevance:** High

Rating Comment:

Moderating Variables

Acquisition Phase: EMD **Contract Type:** FPIF

Program Comments:

S/W Lifecycle: Multiple **Language:** Fortran **Language %:** 100.00%

Application: Database **Project Budget:** 2440000 **Budget Volatility:** Low

Size: 22400 **% New/Modified Code:** 85.00% **Requirements Volatility:** Low

Rebaselining : No **Quality Stds On Contract:** ☐ **Quality Params Tracked :** ☒

Cost Accounting Anomalies: No effort. No +/- three month data

Program Manager Comments:

Cost Data

	Six Months Prior to Rating	Three Months Prior to Rating	Three Months After Rating	Six Months After Rating
Date:	3/30/90	6/30/90	11/30/90	2/28/91
BCWS:	2440	0	0	2440
BCWP:	2416	0	0	2440
ACWP:	2334	0	0	2334
Budget:	2440	0	0	2440
LRE:	2437	0	0	2333

Derived Moderators

Budget Volatility Index: 0 **LRE Volatility Index:** -0.043 **Percent Complete:** 1

BCWS Activity: 0 **BCWP Activity:** 0.00984 **ACWP Activity:** 0

Dependent Variables

Schedule Performance Index: #Error **Cost Performance Index:** #Error

Investigator Comments:

Data point excluded from Complete Data Set due to low activity level. No data for plus/minus three months.

Data Identification

Program Tag: K **RatingTag:** B **Project Tag (WBS#):** 5

Project Description: Requirements, design, code, and test of systems interface CSCI

Rating Information

Rating Date: 9/15/90 **Rating:** 2 **Rating Type:** SCE **Rating Relevance:** High

Rating Comment:

Moderating Variables

Acquisition Phase: **Contract Type:**

Program Comments:

S/W Lifecycle: Multiple **Language:** Fortran **Language %:** 100.00%

Application: Database **Project Budget:** 4236000 **Budget Volatility:** Low

Size: 43200 **% New/Modified Code:** 85.00% **Requirements Volatility:** Low

Rebaselining : No **Quality Stds On Contract:** ☐ **Quality Params Tracked :** ☒

Cost Accounting Anomalies: No +/- three month data

Program Manager Comments:

Cost Data

	Six Months Prior to Rating	Three Months Prior to Rating	Three Months After Rating	Six Months After Rating
Date:	3/30/90	6/30/90	11/30/90	2/28/91
BCWS:	3083	0	0	4236
BCWP:	3019	0	0	4195
ACWP:	2875	0	0	3538
Budget:	4238	0	0	4236
LRE:	4172	0	0	3839

Derived Moderators

Budget Volatility Index: -0.0005 **LRE Volatility Index:** -0.08 **Percent Complete:** 0.9903

BCWS Activity: 0.27219 **BCWP Activity:** 0.28033 **ACWP Activity:** 0.18739

Dependent Variables

Schedule Performance Index: 1.019948 **Cost Performance Index:** 1.77376

Investigator Comments:

No data for plus/minus three month.

Data Identification

Program Tag: K **RatingTag:** B **Project Tag (WBS#):** 6

Project Description: Requirements, design, code, and test of applications CSCI

Rating Information

Rating Date: 9/15/90 **Rating:** 2 **Rating Type:** SCE **Rating Relevance:** High

Rating Comment:

Moderating Variables

Acquisition Phase: **Contract Type:**

Program Comments:

S/W Lifecycle: Multiple **Language:** Fortran **Language %:** 100.00%

Application: Database **Project Budget:** 2683000 **Budget Volatility:** Low

Size: 73200 **% New/Modified Code:** 85.00% **Requirements Volatility:** Low

Rebaselining : No **Quality Stds On Contract:** ☐ **Quality Params Tracked :** ☒

Cost Accounting Anomalies: Negligible effort during this period. No +/- three month data

Program Manager Comments:

Cost Data

	Six Months Prior to Rating	Three Months Prior to Rating	Three Months After Rating	Six Months After Rating
Date:	3/30/90	6/30/90	11/30/90	2/28/91
BCWS:	2666	0	0	2683
BCWP:	2653	0	0	2667
ACWP:	2645	0	0	2649
Budget:	2683	0	0	2683
LRE:	2755	0	0	2667

Derived Moderators

Budget Volatility Index: 0 **LRE Volatility Index:** -0.032 **Percent Complete:** 0.9940

BCWS Activity: 0.00634 **BCWP Activity:** 0.00525 **ACWP Activity:** 0.00151

Dependent Variables

Schedule Performance Index: 0.823529 **Cost Performance Index:** 3.5

Investigator Comments:

Data point excluded from Complete Data Set due to low activity level. No data for plus/minus three months.

Data Identification

Program Tag: K **RatingTag:** B **Project Tag (WBS#):** 7

Project Description: Requirements, design, code, and test of database maintenance CSCI

Rating Information

Rating Date: 9/15/90 **Rating:** 2 **Rating Type:** SCE **Rating Relevance:** High

Rating Comment:

Moderating Variables

Acquisition Phase: **Contract Type:**

Program Comments:

S/W Lifecycle: Multiple **Language:** Fortran **Language %:** 100.00%

Application: Database **Project Budget:** 2666000 **Budget Volatility:** Low

Size: 25700 **% New/Modified Code:** 85.00% **Requirements Volatility:** Low

Rebaselining : No **Quality Stds On Contract:** ☐ **Quality Params Tracked :** ☒

Cost Accounting Anomalies: Negligible effort during this period. No +/- three month data

Program Manager Comments:

Cost Data

	Six Months Prior to Rating	Three Months Prior to Rating	Three Months After Rating	Six Months After Rating
Date:	3/30/90	6/30/90	11/30/90	2/28/91
BCWS:	2650	0	0	2666
BCWP:	2650	0	0	2666
ACWP:	2866	0	0	2870
Budget:	2667	0	0	2666
LRE:	2874	0	0	2870

Derived Moderators

Budget Volatility Index: -0.0004 **LRE Volatility Index:** -0.001 **Percent Complete:** 1

BCWS Activity: 0.00600 **BCWP Activity:** 0.00600 **ACWP Activity:** 0.00139

Dependent Variables

Schedule Performance Index: 1 **Cost Performance Index:** 4

Investigator Comments:

Data point excluded from Complete Data Set due to low activity level. No data for plus/minus three months.

Data Identification

Program Tag: K   RatingTag: B   Project Tag (WBS#): 8

Project Description: Requirements, design, code, and test of database support CSC

Rating Information

Rating Date: 9/15/90   Rating: 2   Rating Type: SCE   Rating Relevance: High

Rating Comment:

Moderating Variables

Acquisition Phase: Contract Type:

Program Comments:

S/W Lifecycle: Multiple   Language: Fortran   Language %: 100.00%

Application: Database   Project Budget: 1181000   Budget Volatility: Low

Size: 14200   % New/Modified Code: 85.00%   Requirements Volatility: Low

Rebaselining : No   Quality Stds On Contract: ☐   Quality Params Tracked : ☒

Cost Accounting Anomalies: Negligible effort during this period. No +/- three month data

Program Manager Comments:

Cost Data

	Six Months Prior to Rating	Three Months Prior to Rating	Three Months After Rating	Six Months After Rating
Date:	3/30/90	6/30/90	11/30/90	2/28/91
BCWS:	1175	0	0	1181
BCWP:	1175	0	0	1181
ACWP:	1266	0	0	1269
Budget:	1181	0	0	1181
LRE:	1277	0	0	1269

Derived Moderators

Budget Volatility Index: 0   LRE Volatility Index: -0.006   Percent Complete: 1

BCWS Activity: 0.00508   BCWP Activity: 0.00508   ACWP Activity: 0.00236

Dependent Variables

Schedule Performance Index: 1   Cost Performance Index: 2

Investigator Comments:

Data point excluded from Complete Data Set due to low activity level. No data for plus/minus three months.

Data Identification

Program Tag: K RatingTag: B Project Tag (WBS#): 9

Project Description: Software integration activities.

Rating Information

Rating Date: 9/15/90 Rating: 2 Rating Type: SCE Rating Relevance: High

Rating Comment:

Moderating Variables

Acquisition Phase: Contract Type:

Program Comments:

S/W Lifecycle: Test/Integration Language: Fortran Language %: 100.00%

Application: Database Project Budget: 6874000 Budget Volatility: Low

Size: 0 % New/Modified Code: 85.00% Requirements Volatility: Low

Rebaselining : No Quality Stds On Contract: ☐ Quality Params Tracked : ☒

Cost Accounting Anomalies: No +/- three month data

Program Manager Comments:

Cost Data

	Six Months Prior to Rating	Three Months Prior to Rating	Three Months After Rating	Six Months After Rating
Date:	3/30/90	6/30/90	11/30/90	2/28/91
BCWS:	4564	0	0	6486
BCWP:	4426	0	0	6486
ACWP:	7084	0	0	9461
Budget:	5821	0	0	6874
LRE:	7384	0	0	10014

Derived Moderators

Budget Volatility Index: 0.1809 LRE Volatility Index: 0.3562 Percent Complete: 0.9436

BCWS Activity: 0.29633 BCWP Activity: 0.31761 ACWP Activity: 0.25124

Dependent Variables

Schedule Performance Index: 1.071800 Cost Performance Index: 0.86664

Investigator Comments:

No data for plus/minus three month.

Data Identification

Program Tag: L **RatingTag:** A **Project Tag (WBS#):** 1

Project Description: Generates all sytem design requirements (logic & algorithms) and software to support technology item being developed

Rating Information

Rating Date: 5/15/92 **Rating:** 2 **Rating Type:** SPA (EXT) **Rating Relevance:** Low

Rating Comment: Conducted in accordance with an SEI-licensed vendor agreement between *vendor* and SEI

Moderating Variables

Acquisition Phase: Concept Exploration **Contract Type:** CPI

Program Comments: 85% software, 15% hardware. Program partially terminated after technology demonstrated.

S/W Lifecycle: Multiple **Language:** Ada **Language %:** 100.00%

Application: Avionics **Project Budget:** 2726000 **Budget Volatility:** Low

Size: 76636 **% New/Modified Code:** 100.00% **Requirements Volatility:** Med

Rebaselining : No **Quality Stds On Contract:** **Quality Params Tracked :**

Cost Accounting Anomalies: No agreement between Govt and Contractor on Estimate to Complete. Contractor may have tried to "get well" on options. Contractor may have taken earned value early.

Program Manager Comments: Requirements changes due to interfaces with associate contractor. Overruns covered by termination agreement. Language was early Ada (non-validated compiler). Contractor cited too much documentation as reason for overrun.

Cost Data

	Six Months Prior to Rating	Three Months Prior to Rating	Three Months After Rating	Six Months After Rating
Date:	12/30/91	3/30/92	8/30/92	11/30/92
BCWS:	2246	2335	2716	2739
BCWP:	2025	2203	2309	2369
ACWP:	2937	3112	3296	3367
Budget:	2716	2726	2726	2726
LRE:	3222	3226	3226	3226

Derived Moderators

Budget Volatility Index: 0.00368 **LRE Volatility Index:** 0.0012 **Percent Complete:** 0.8690

BCWS Activity: 0.17999 **BCWP Activity:** 0.14521 **ACWP Activity:** 0.12771

Dependent Variables

Schedule Performance Index: 0.697769 **Cost Performance Index:** 0.8

Investigator Comments:

Data Identification

Program Tag: N **RatingTag:** A **Project Tag (WBS#):** 1

Project Description: Modify existing software for new configuration

Rating Information

Rating Date: 10/15/92 **Rating:** 2 **Rating Type:** SPA (INT) **Rating Relevance:** High

Rating Comment: Performed by a former SEI employee: "borderline"

Moderating Variables

Acquisition Phase: EMD **Contract Type:** CPI

Program Comments:

S/W Lifecycle: Multiple-Early **Language:** Fortran **Language %:** 90.00%

Application: Command & Co **Project Budget:** 2230000 **Budget Volatility:** Low

Size: 550000 **% New/Modified Code:** 80.00% **Requirements Volatility:** Low

Rebaselining : No **Quality Stds On Contract:** ☐ **Quality Params Tracked :** ☒

Cost Accounting Anomalies: Increasing baseline reflected througth ECPs

Program Manager Comments:

Cost Data

	Six Months Prior to Rating	Three Months Prior to Rating	Three Months After Rating	Six Months After Rating
Date:	5/30/92	8/30/92	1/30/93	4/30/93
BCWS:	0	530	1688	2138
BCWP:	0	375	1483	2080
ACWP:	0	300	1138	1812
Budget:	2227	2227	2226	2230
LRE:	2227	2227	2172	2012

Derived Moderators

Budget Volatility Index: 0.00135 **LRE Volatility Index:** -0.097 **Percent Complete:** 0.9327

BCWS Activity: 1 **BCWP Activity:** 1 **ACWP Activity:** 1

Dependent Variables

Schedule Performance Index: 0.972872 **Cost Performance Index:** 1.14790

Investigator Comments:

Data Identification

Program Tag: N **RatingTag:** B **Project Tag (WBS#):** 1

Project Description: Modify existing software for new configuration

Rating Information

Rating Date: 9/15/93 **Rating:** 1 **Rating Type:** SCE **Rating Relevance:** High

Rating Comment: Contractor stated rating of level 1 due to QA on another program. Rating information provided by contractor with Program Office permission.

Moderating Variables

Acquisition Phase: EMD **Contract Type:** CPI

Program Comments:

S/W Lifecycle: Test/Integration **Language:** Fortran **Language %:** 90.00%

Application: Command & Co **Project Budget:** 2268000 **Budget Volatility:** Low

Size: 550000 **% New/Modified Code:** 80.00% **Requirements Volatility:** Low

Rebaselining : No **Quality Stds On Contract:** ☐ **Quality Params Tracked :** ☒

Cost Accounting Anomalies: Increasing baseline reflected througth ECPs

Program Manager Comments:

Cost Data

	Six Months Prior to Rating	Three Months Prior to Rating	Three Months After Rating	Six Months After Rating
Date:	3/30/93	6/30/93	11/30/93	2/28/94
BCWS:	2025	2199	2268	2268
BCWP:	1947	2190	2257	2257
ACWP:	1694	1862	1974	2096
Budget:	2230	2268	2268	2268
LRE:	2176	2076	1995	2222

Derived Moderators

Budget Volatility Index: 0.01704 **LRE Volatility Index:** 0.0211 **Percent Complete:** 0.9951

BCWS Activity: 0.10714 **BCWP Activity:** 0.13735 **ACWP Activity:** 0.19179

Dependent Variables

Schedule Performance Index: 1.275720 **Cost Performance Index:** 0.77114

Investigator Comments:

Data Identification

Program Tag: O **RatingTag:** A **Project Tag (WBS#):** 1

Project Description: Design, code, test, integration of all software for entire system consisting of 3 major components

Rating Information

Rating Date: 2/15/94 **Rating:** 1 **Rating Type:** SPA (INT) **Rating Relevance:** High

Rating Comment:

Moderating Variables

Acquisition Phase: EMD **Contract Type:** CPAF

Program Comments:

S/W Lifecycle: Design/Code **Language:** Ada **Language %:** 100.00%

Application: Simulation **Project Budget:** 3153000 **Budget Volatility:** Low

Size: 130000 **% New/Modified Code:** 100.00% **Requirements Volatility:** Med

Rebaselining : No **Quality Stds On Contract:** [X] **Quality Params Tracked :** [X]

Cost Accounting Anomalies:

Program Manager Comments: Company does not have domain expertise. ECPs drivers of cost growth.

Cost Data

	Six Months Prior to Rating	Three Months Prior to Rating	Three Months After Rating	Six Months After Rating
Date:	8/30/93	11/30/93	4/30/94	7/30/94
BCWS:	1561	1874	2767	2943
BCWP:	1431	1609	2077	2192
ACWP:	2448	3327	4725	5669
Budget:	2889	2900	2950	3153
LRE:	4392	5378	6703	6980

Derived Moderators

Budget Volatility Index: 0.09138 **LRE Volatility Index:** 0.5893 **Percent Complete:** 0.6952

BCWS Activity: 0.46959 **BCWP Activity:** 0.34717 **ACWP Activity:** 0.56818

Dependent Variables

Schedule Performance Index: 0.550651 **Cost Performance Index:** 0.23626

Investigator Comments:

Appendix C: Data Supporting the Analysis of the Complete Data Set

This appendix contains the complete set of plots, tables, and calculations supporting sections 5.2. The plots and the Kruskal-Wallis tables for nonparametric analysis of variance are from the Statistix 4.0 computer program. The multiple comparison calculations were performed using Mathcad 4.0. Note that we did not abridge the data as we transcribed it from these computer programs into the report, and thus the number of digits reported in each calculation are not necessarily significant.

1. Scatter Plots of CPI and SPI

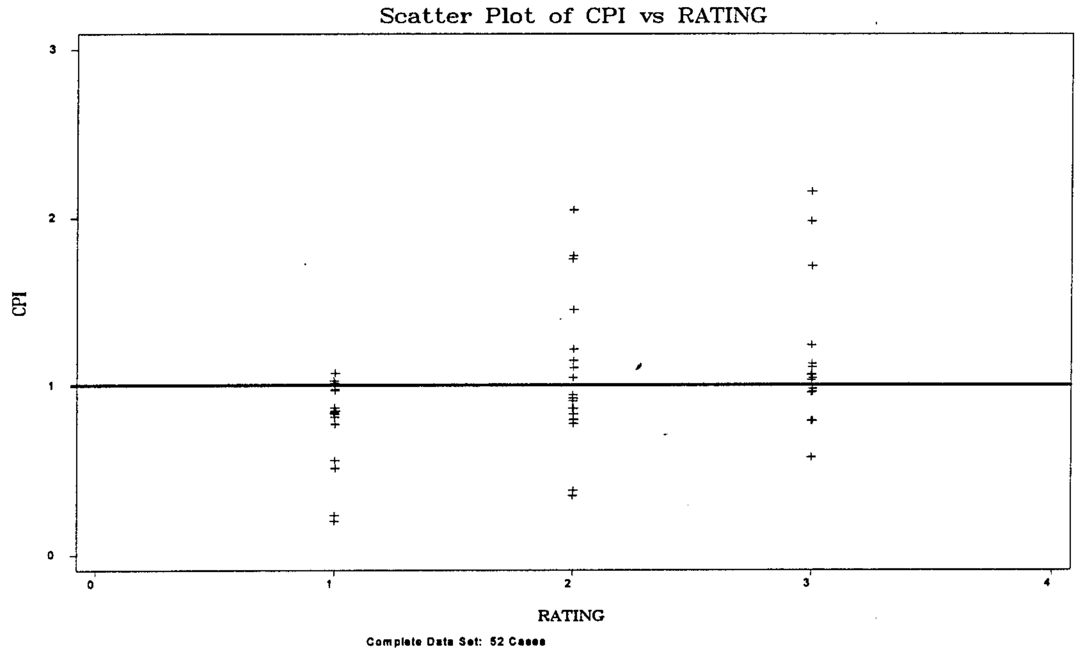

Figure C-1 Scatter Plot of CPI versus Rating for the Complete Data Set

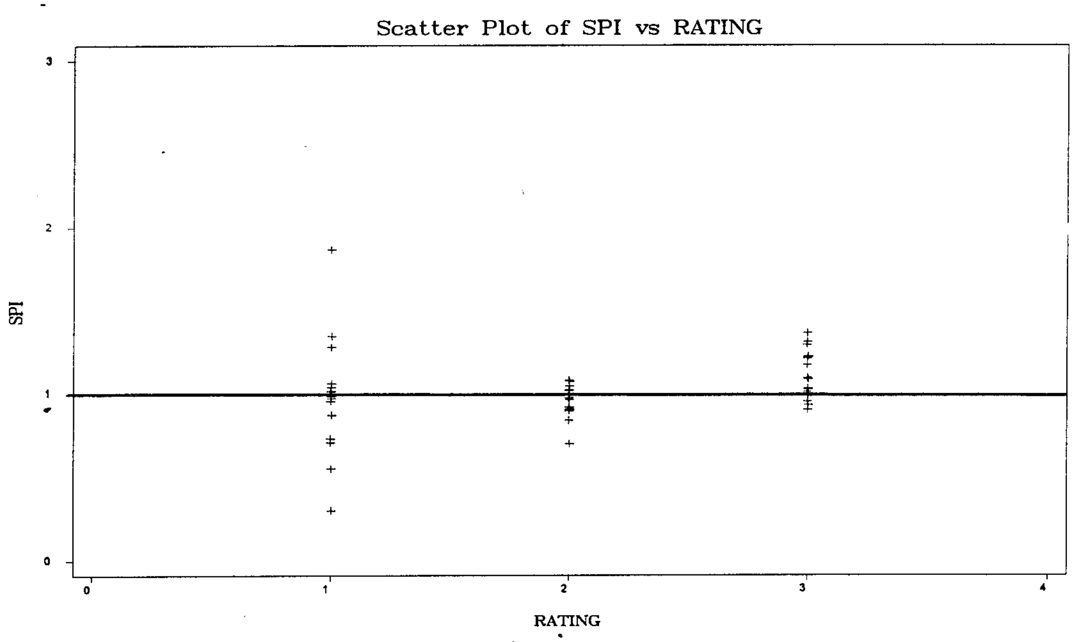

Figure C-2 Scatter Plot of SPI versus Rating for the Complete Data Set

2. Histogram of the frequency density for each rating level

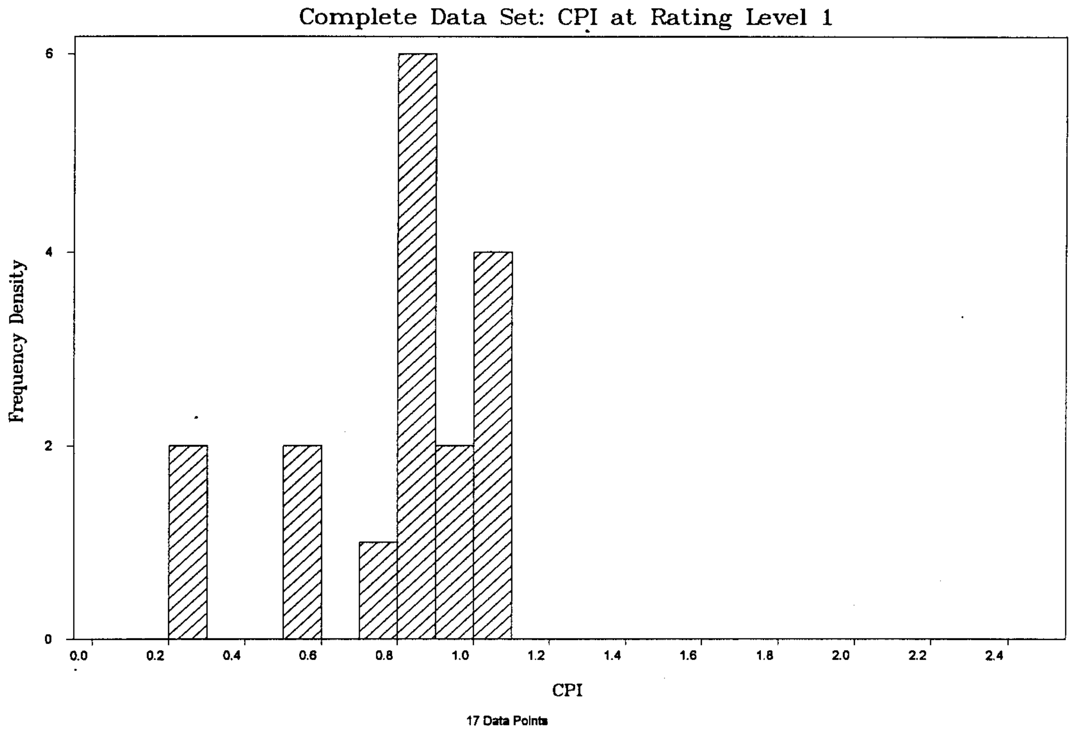

Figure C-3 Histogram of CPI at Rating Level One for the Complete Data Set

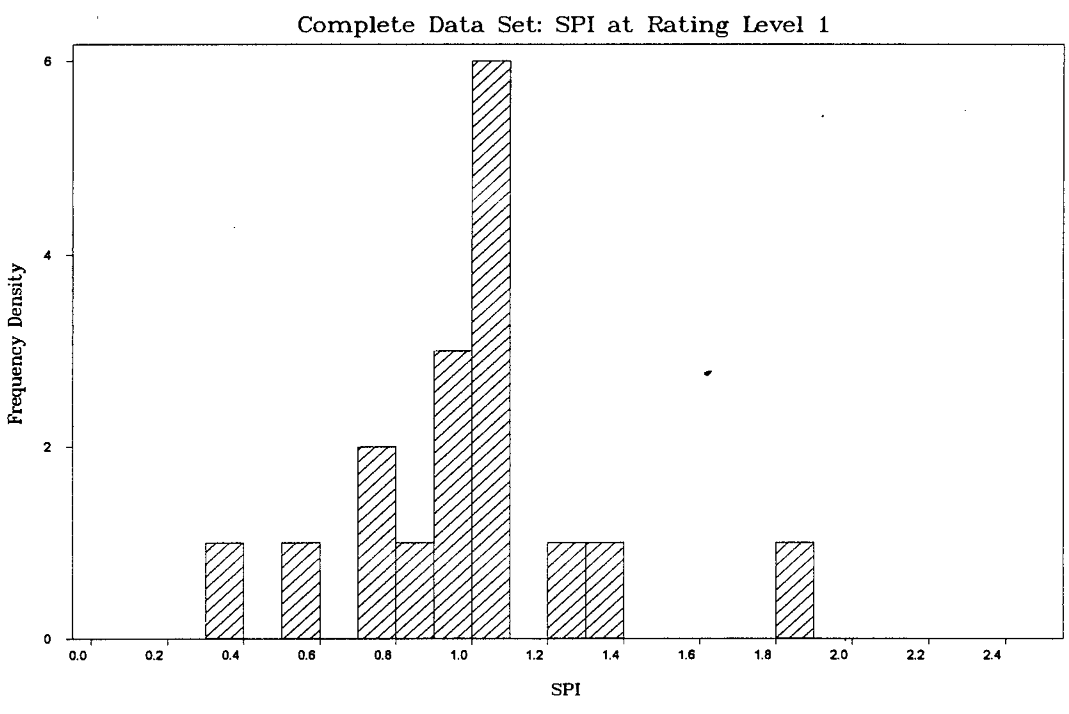

Figure C-4 Histogram for SPI at Rating Level One for the Complete Data Set

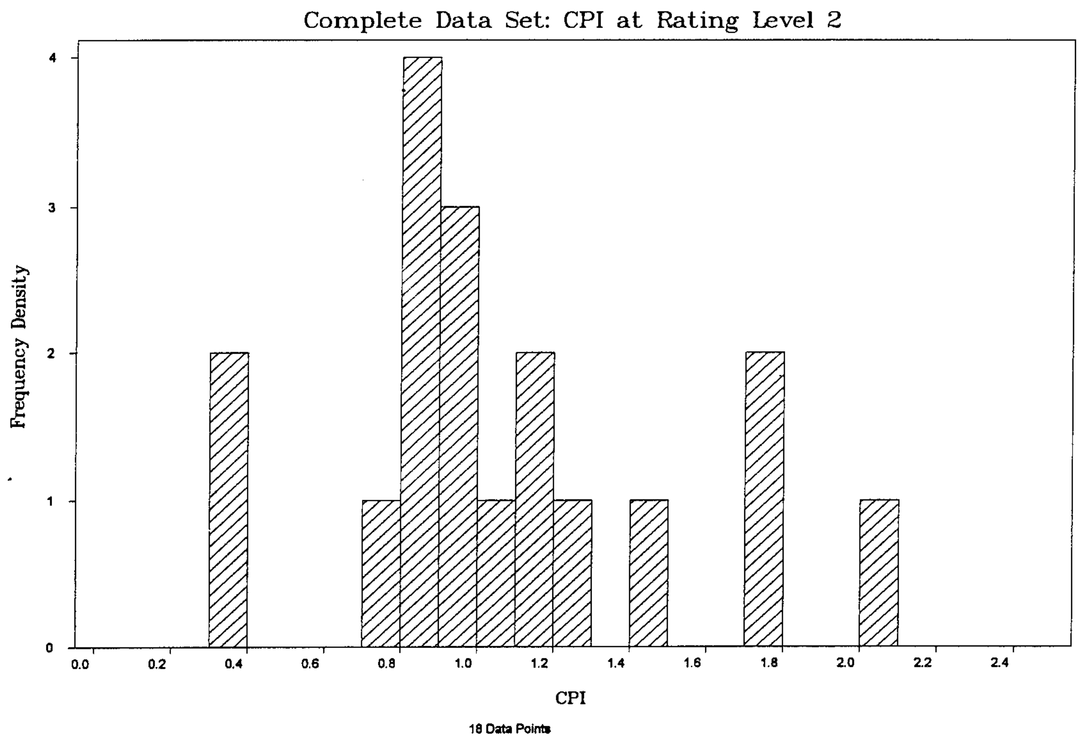

Figure C-5 Histogram for CPI at Rating Level Two for the Complete Data Set

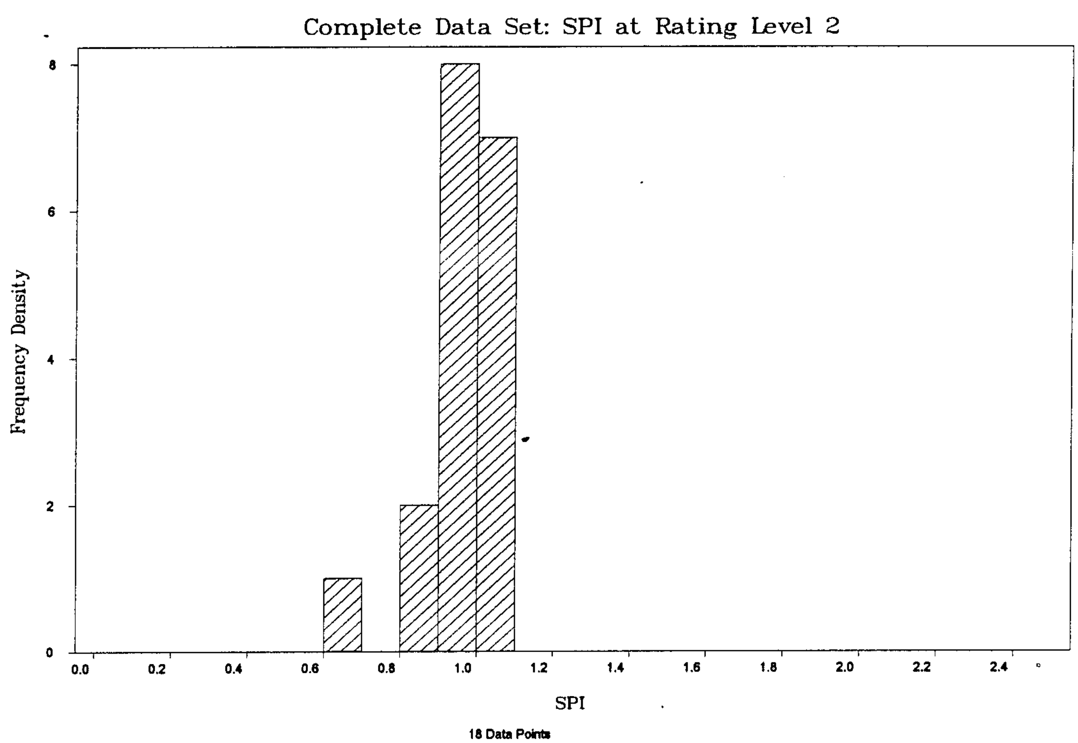

Figure C-6 Histogram for SPI at Rating Level Two for the Complete Data Set

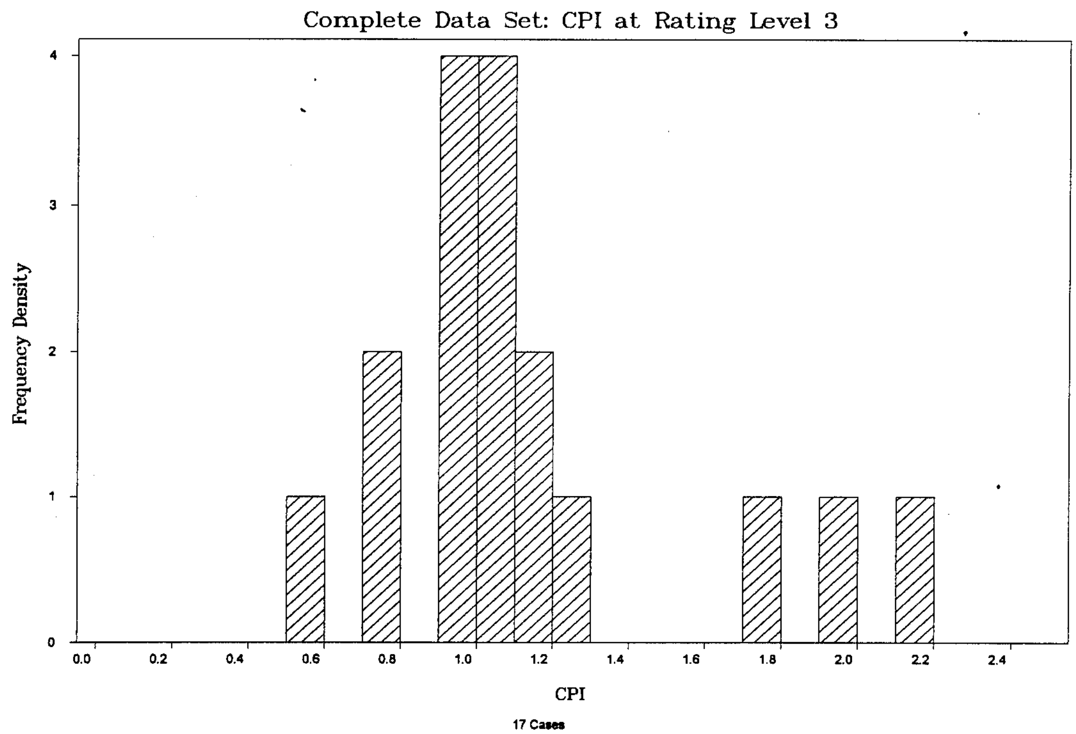

Figure C-7 Histogram for CPI at Rating Level Three for the Complete Data Set

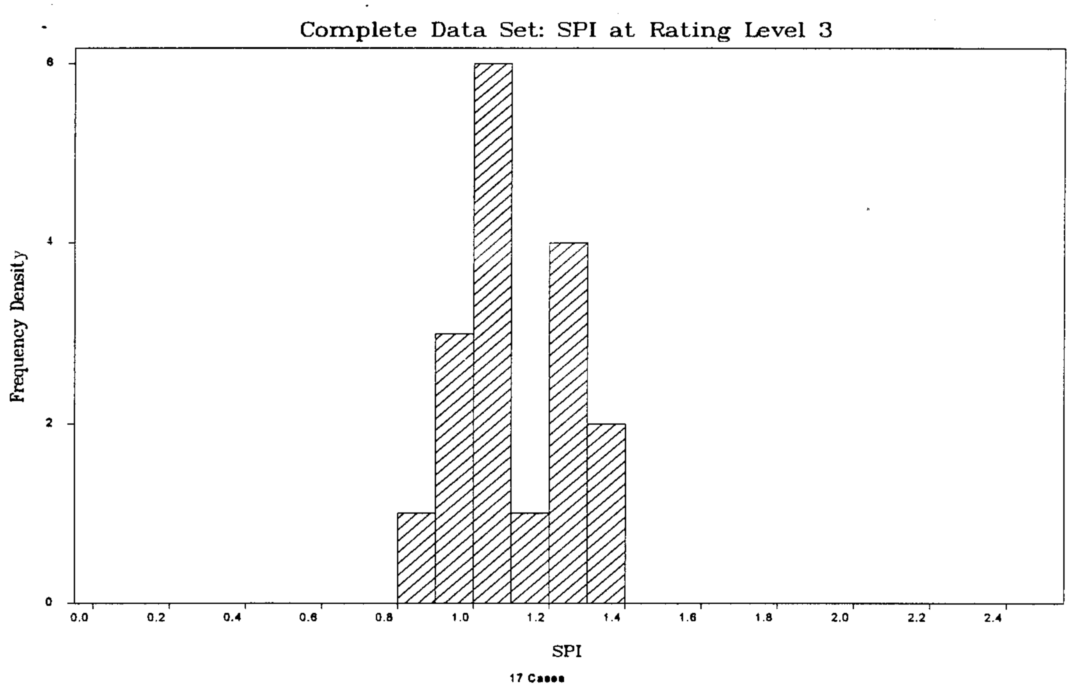

Figure C-8 Histogram for SPI at Rating Level Three for the Complete Data Set

3. Wilk-Shapiro evaluation of normality at each level

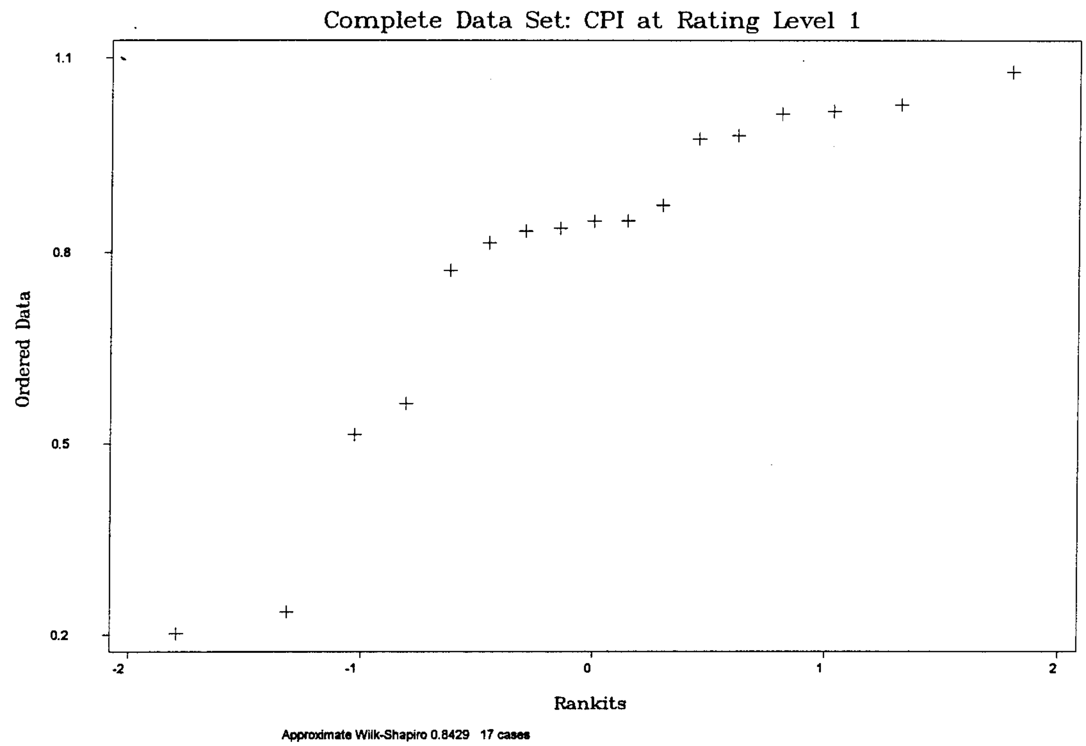

Figure C-9 Wilk-Shapiro Plot for CPI at Rating Level One for the Complete Data Set

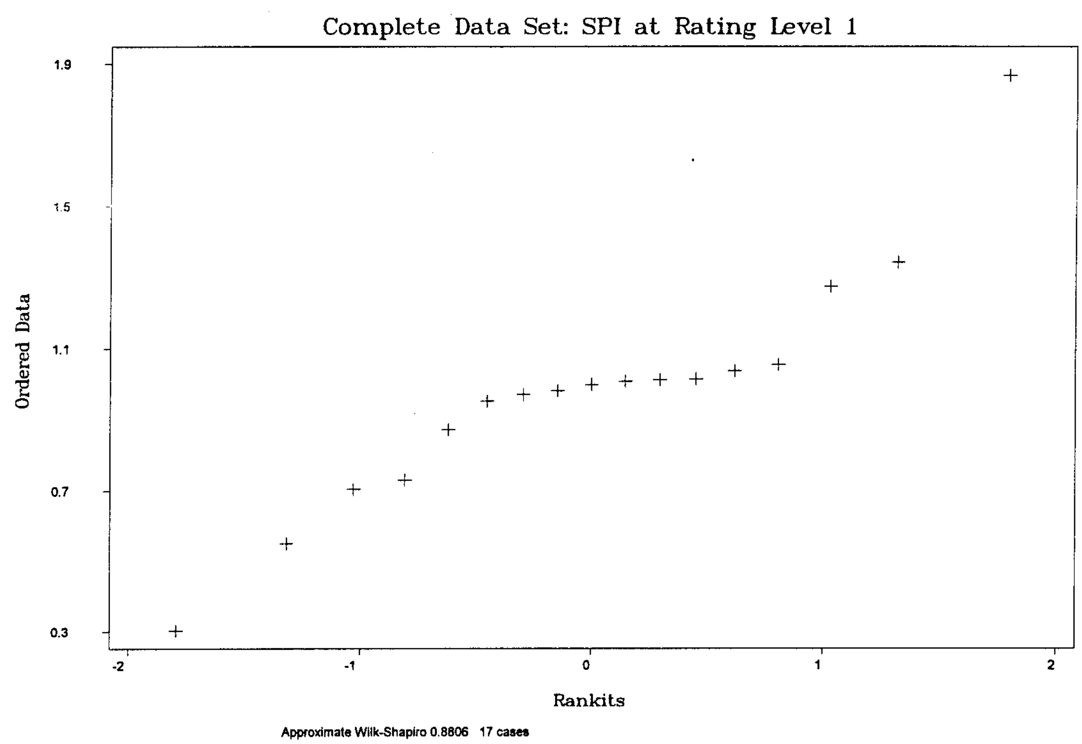

Figure C-10 Wilk-Shapiro Plot for SPI at Rating Level One for the Complete Data Set

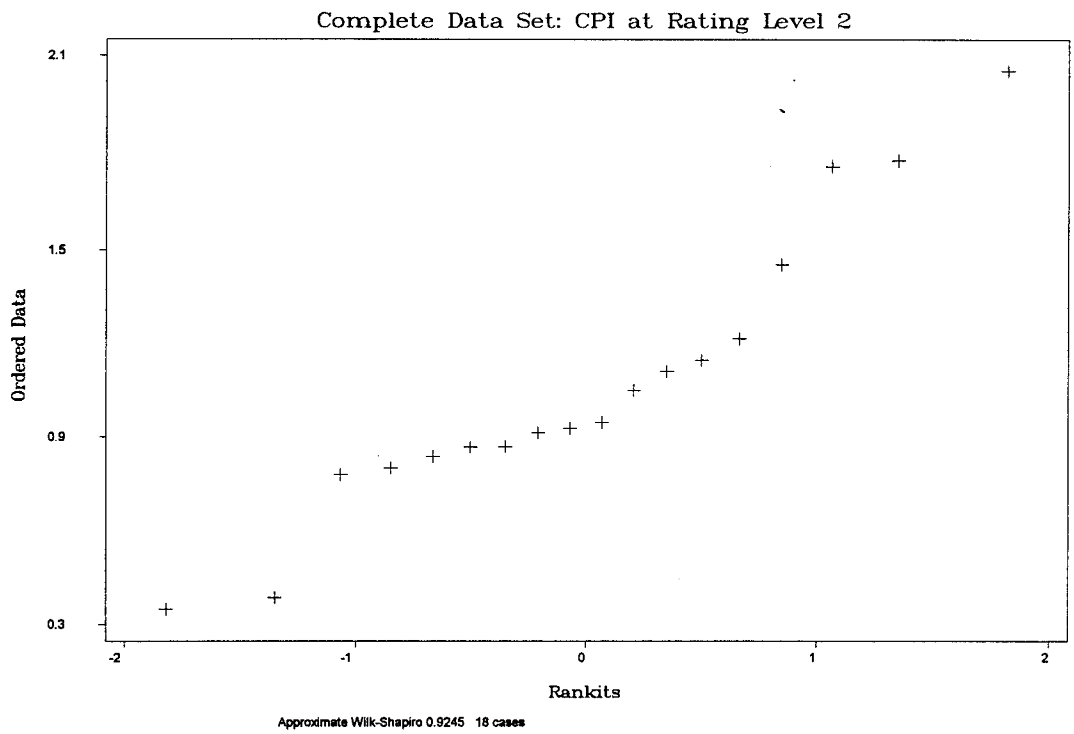

Figure C-11 Wilk-Shapiro Plot for CPI at Rating Level Two for the Complete Data Set

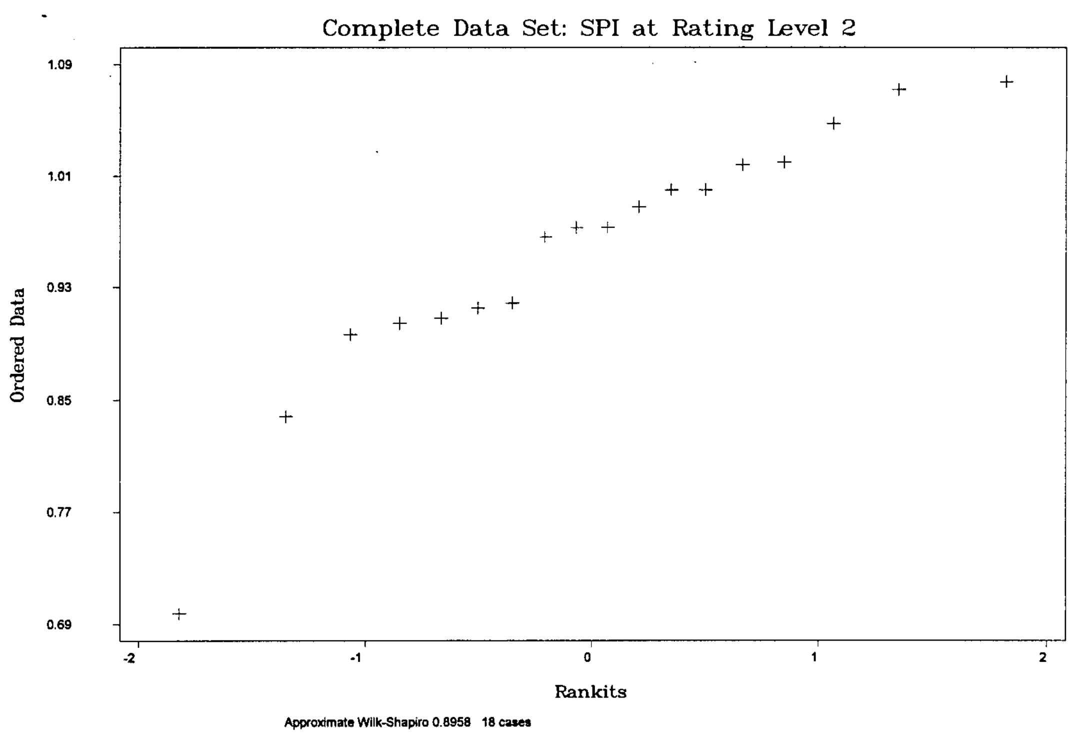

Figure C-12 Wilk-Shapiro Plot for SPI at Rating Level Two for the Complete Data Set

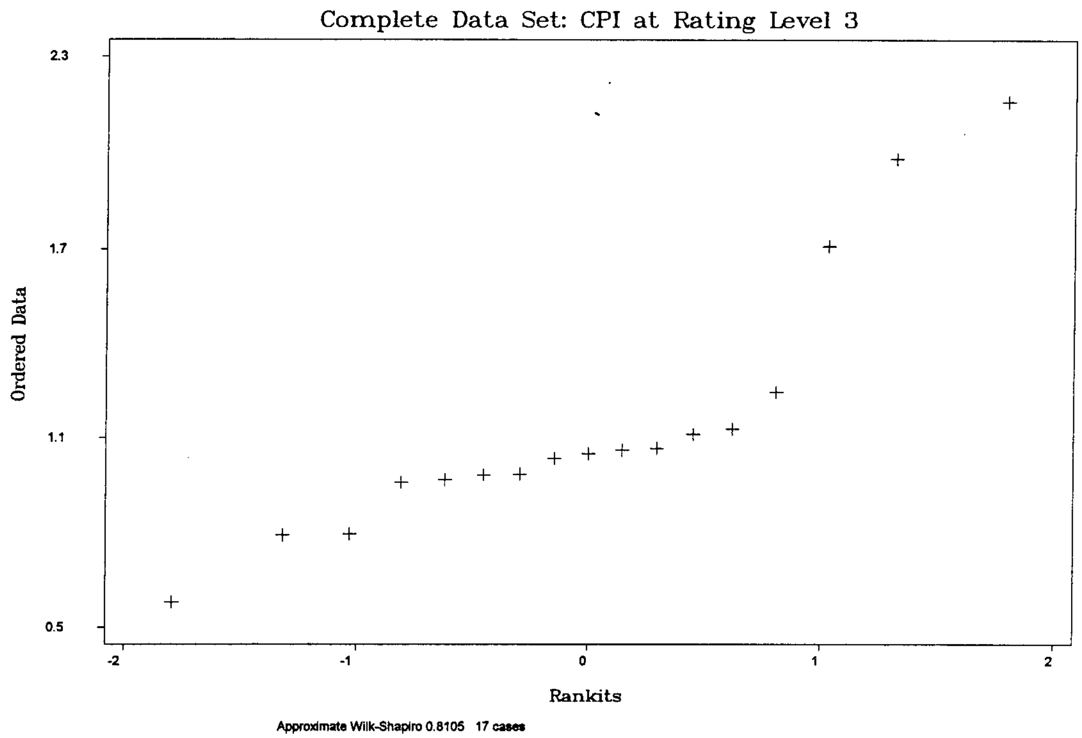

Figure C-13 Wilk-Shapiro Plot for CPI at Rating Level Three for the Complete Data Set

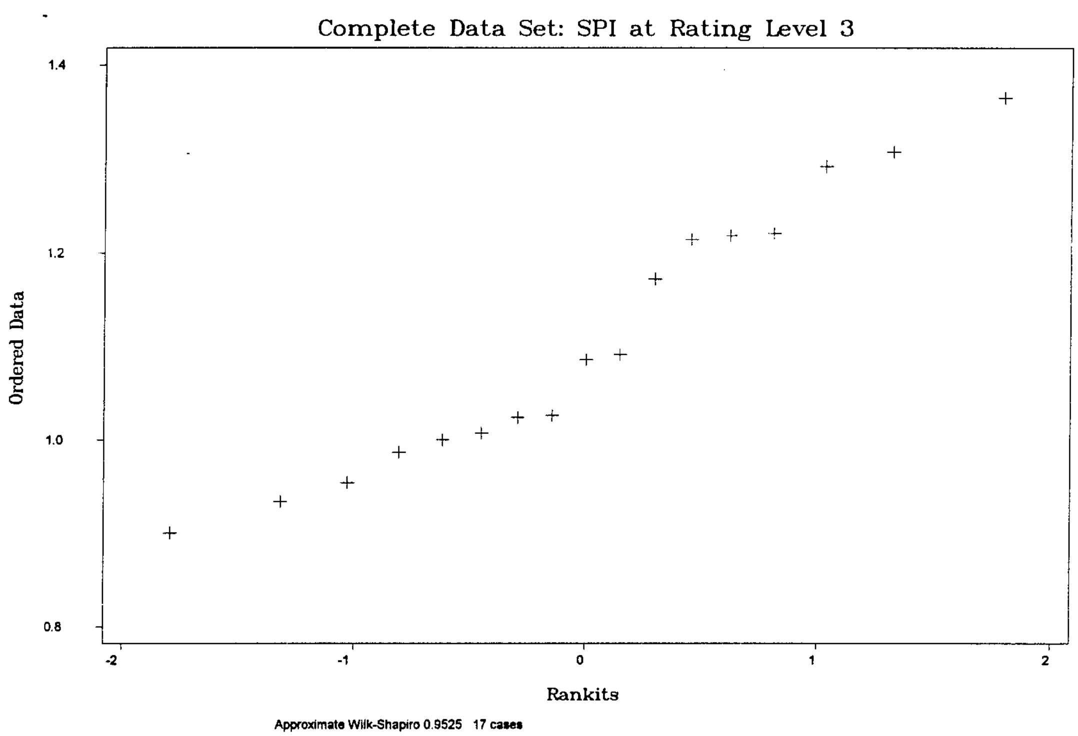

Figure C-14 Wilk-Shapiro Plot for SPI at Rating Level Three for the Complete Data Set

4. Box and Whiskers Plots of CPI and SPI

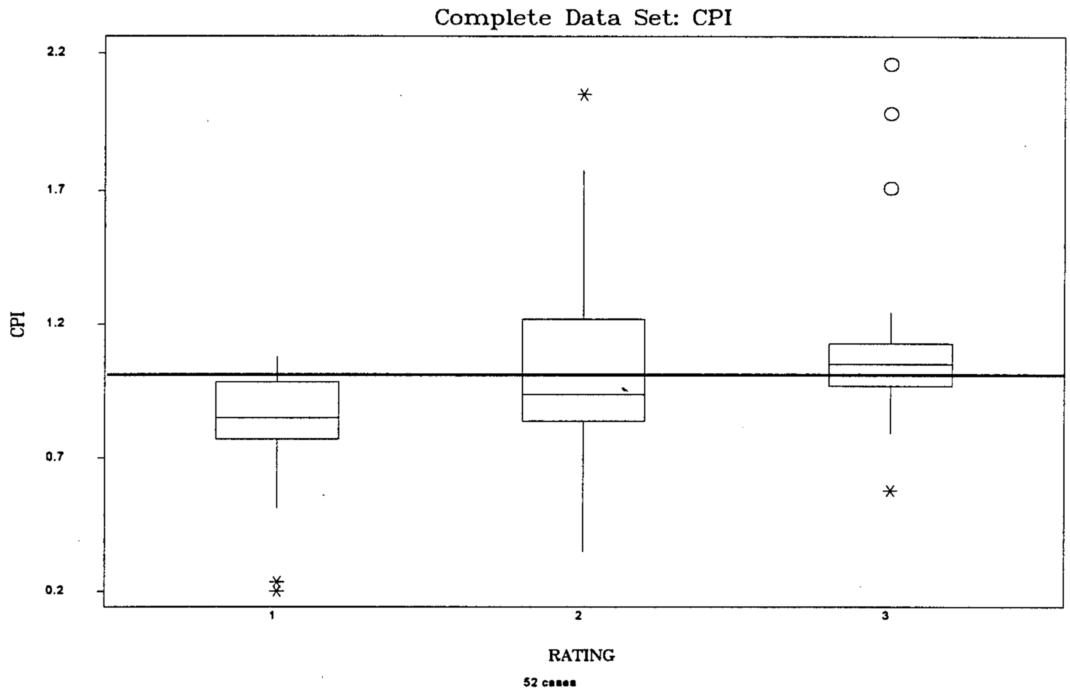

Figure C-15 Box and Whiskers Plot for CPI for the Complete Data Set

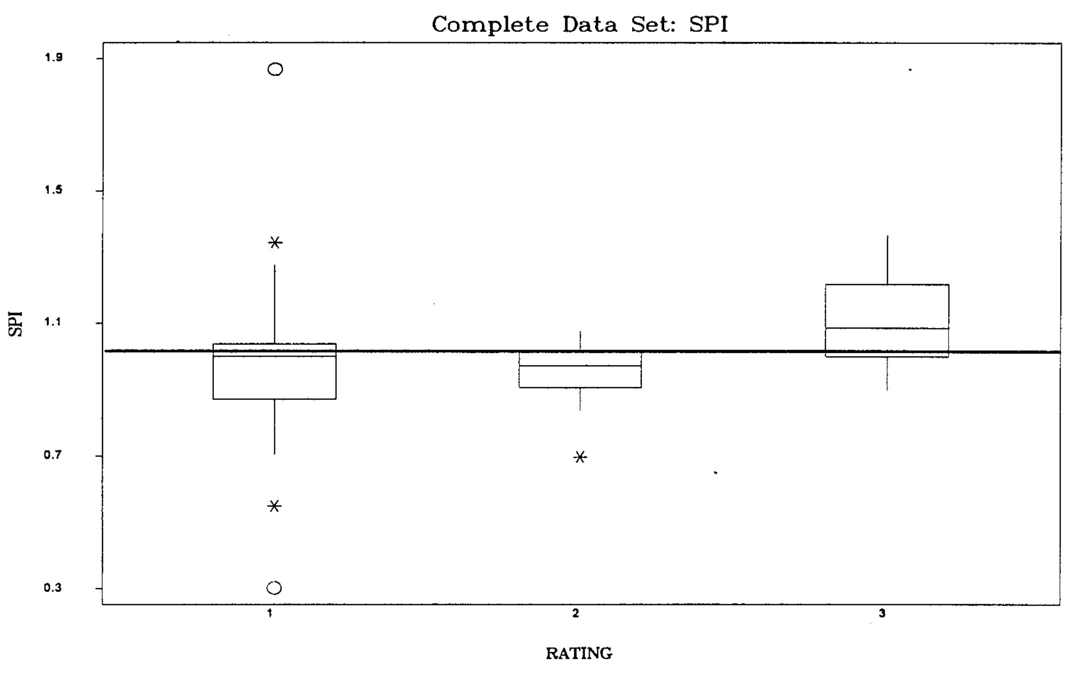

Figure C-16 Box and Whiskers Plot for SPI for the Complete Data Set

5. Kruskal-Wallis Tests, and Multiple Comparison Tests

Table C-1
Kruskal-Wallis for CPI for the Complete Data Set

```
KRUSKAL-WALLIS ONE-WAY NONPARAMETRIC AOV FOR CPI BY RATING

           MEAN   SAMPLE
RATING     RANK    SIZE
  1        18.3    17
  2        28.2    18
  3        32.9    17
TOTAL      26.5    52

KRUSKAL-WALLIS STATISTIC                       8.2319
P-VALUE, USING CHI-SQUARED APPROXIMATION       0.0163
```

Table C-2
Multiple Comparison Matrix for CPI for the Complete Data Set

			Rating		
Rating	n	Mean Rank	1	2	3
1	17	18.3		--	--
2	18	28.2	0.5		--
3	17	32.9	5.067	-4.7	

K-W Statistic of 8.2319, P=0.0163

Table C-3
Kruskal-Wallis for SPI for the Complete Data Set

```
KRUSKAL-WALLIS ONE-WAY NONPARAMETRIC AOV FOR SPI BY RATING

           MEAN   SAMPLE
RATING     RANK    SIZE
  1        24.2    17
  2        20.8    18
  3        34.9    17
TOTAL      26.5    52

KRUSKAL-WALLIS STATISTIC                       8.1238
P-VALUE, USING CHI-SQUARED APPROXIMATION       0.0172
```

Table C-4
Multiple Comparison Matrix for SPI for the Complete Data Set

			Rating		
Rating	n	Mean Rank	1	2	3
1	17	24.2		--	--
2	18	20.8	-6.0		--
3	17	34.9	1.167	4.7	

K-W Statistic of 8.1238, P=0.0172

6. Descriptive Statistics of the Complete Data Set

Table C-5

Descriptive Statistics for the Complete Data Set

	Rating=1 CPI	Rating=1 SPI	Rating=2 CPI	Rating=2 SPI	Rating=3 CPI	Rating=3 SPI
N	17	17	18	18	17	17
Mean	0.7909	0.9816	1.0685	0.9562	1.1537	1.1059
Std Dev	0.2639	0.3366	0.4502	0.0915	0.4165	0.1433
Min	0.2019	0.3028	0.3496	0.6978	0.5808	0.8998
Median	0.8493	1.0000	0.9365	0.9727	1.0498	1.0864
Max	1.0788	1.8676	2.0506	1.0774	2.1602	1.3652
MAD	0.1325	0.0560	0.1661	0.0559	0.0825	0.1280
Skew	-1.1674	0.5698	0.6062	-1.1639	1.294	0.3124

7. *Multiple Comparison Calculations for the Complete Data Set*

The Kruskal-Wallis H Test Statistic

$$H := \frac{12}{N\cdot(N+1)}\cdot\sum_{j=1}^{k}\frac{\left[R_j - \frac{n_j\cdot(N+1)}{2}\right]^2}{n_j}$$

Multiple Comparison Inequality

$$\left|\overline{R_i} - \overline{R_j}\right| \le z\cdot\sqrt{\frac{N\cdot(N+1)}{12}\cdot\left(\frac{1}{n_i}+\frac{1}{n_j}\right)}$$

where
k = number of samples
R_i, R_j = sum of the ranks in the ith and jth sample, respectively
n_i, n_j = number of observations in the ith and jth sample, respectively
N = total number of observations

Mean Rank of Ratings 1, 2, & 3 for CPI
$R_{cpi} := (18.3\ \ 28.2\ \ 32.9)$

Mean Rank of Ratings 1, 2, & 3 for SPI
$R_{spi} := (24.2\ \ 20.8\ \ 34.9)$

Number of Observations in Ratings 1, 2, & 3
$n := (17\ \ 18\ \ 17)$

Calculation of N $\quad N := n_{1,1} + n_{1,2} + n_{1,3} \quad N = 52$

Critical z Value - Corresponds to a .2 level of significance and provides at least a 80% level of confidence.
$z := 1.834$

<u>**Multiple Comparisons**</u> - "which populations differ from which others"
- If the difference in mean rank (DMR) between 2 samples is greater than the right-hand side (RHS) of the multiple comparison inequality above, there is a statistically significant difference between the 2 samples.
- The direction of the significant difference is determined by noting for each pair which sample has the larger sample mean.

Comparision of Rating 1 and 2

$DMR_{cpi} := \left|R_{cpi_{1,1}} - R_{cpi_{1,2}}\right| \quad DMR_{cpi} = 9.9$

$DMR_{spi} := \left|R_{spi_{1,1}} - R_{spi_{1,2}}\right| \quad DMR_{spi} = 3.4$

$RHS := z\cdot\sqrt{\frac{N\cdot(N+1)}{12}\cdot\left(\frac{1}{n_{1,1}}+\frac{1}{n_{1,2}}\right)} \quad RHS = 9.4$

For CPI $DMR_{cpi} - RHS = 0.5$ — CPI at Rating 2 > CPI at Rating 1 <==

For SPI $DMR_{spi} - RHS = -6$ — no significant difference <==

Comparision of Rating 2 and 3

$DMR_{cpi} := \left|R_{cpi_{1,2}} - R_{cpi_{1,3}}\right| \quad DMR_{cpi} = 4.7$

$DMR_{spi} := \left|R_{spi_{1,2}} - R_{spi_{1,3}}\right| \quad DMR_{spi} = 14.1$

$RHS := z\cdot\sqrt{\frac{N\cdot(N+1)}{12}\cdot\left(\frac{1}{n_{1,2}}+\frac{1}{n_{1,3}}\right)} \quad RHS = 9.4$

For CPI $DMR_{cpi} - RHS = -4.7$ — no significant difference <==

For SPI $DMR_{spi} - RHS = 4.7$ — SPI at Rating 3 > SPI at Rating 2 <==

Comparision of Rating 1 and 3

$DMR_{cpi} := \left|R_{cpi_{1,1}} - R_{cpi_{1,3}}\right| \quad DMR_{cpi} = 14.6$

$DMR_{spi} := \left|R_{spi_{1,1}} - R_{spi_{1,3}}\right| \quad DMR_{spi} = 10.7$

$RHS := z\cdot\sqrt{\frac{N\cdot(N+1)}{12}\cdot\left(\frac{1}{n_{1,1}}+\frac{1}{n_{1,3}}\right)} \quad RHS = 9.533$

For CPI $DMR_{cpi} - RHS = 5.067$ — CPI at Rating 3 > CPI at Rating 1 <==

For SPI $DMR_{spi} - RHS = 1.167$ — SPI at Rating 3 > SPI at Rating 1 <==

Figure C-17 Calculations for Multiple Comparison Test for the Complete Data Set

Appendix D: Data Supporting the Analysis of the Moderator "Rating Relevance"

This appendix contains the complete set of plots, tables, and calculations supporting sections 5.2.1. The plots and the Kruskal-Wallis tables for nonparametric analysis of variance are from the Statistix 4.0 computer program. The multiple comparison calculations were performed using Mathcad 4.0. Note that we did not abridge the data as we transcribed it from these computer programs into the report, and thus the number of digits reported in each calculation are not necessarily significant.

For each project, the validity of the correlation between the CMM rating and project cost/schedule performance depends upon the associative relevance of project under consideration. Four scenarios define the four degrees of rating-to-project associative relevance:

1. Very High Rating-to-Project Relevance--the project under consideration was itself rated using the CMM rating process.
2. High Rating-to-Project Relevance--the project under consideration was one project of several used in obtaining the CMM rating for the organization.
3. Medium Rating-to-Project Relevance--the project was not used to establish the CMM rating, but the personnel which participated in the project were responsible for other projects which were evaluated in the CMM rating.
4. Low Rating-to-Project Relevance--neither the project, nor the personnel responsible for the project were used to obtain the organization's CMM rating.

1. Scatter Plots of CPI and SPI

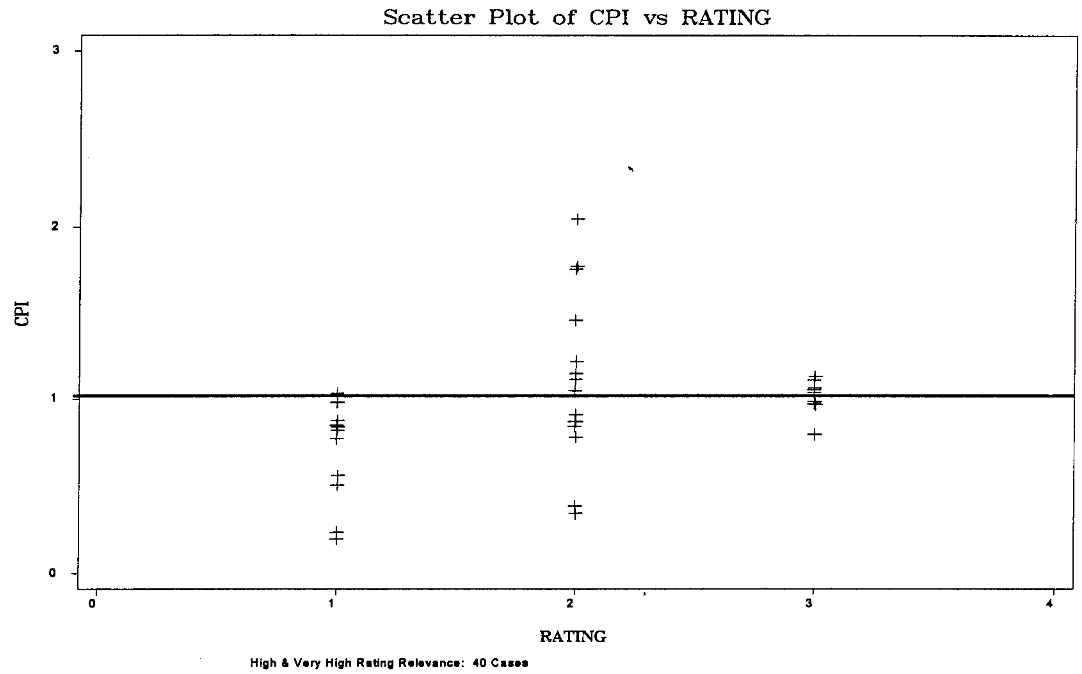

Figure D-1 Scatter Plot of CPI versus Rating for High and Very High Rating Relevance

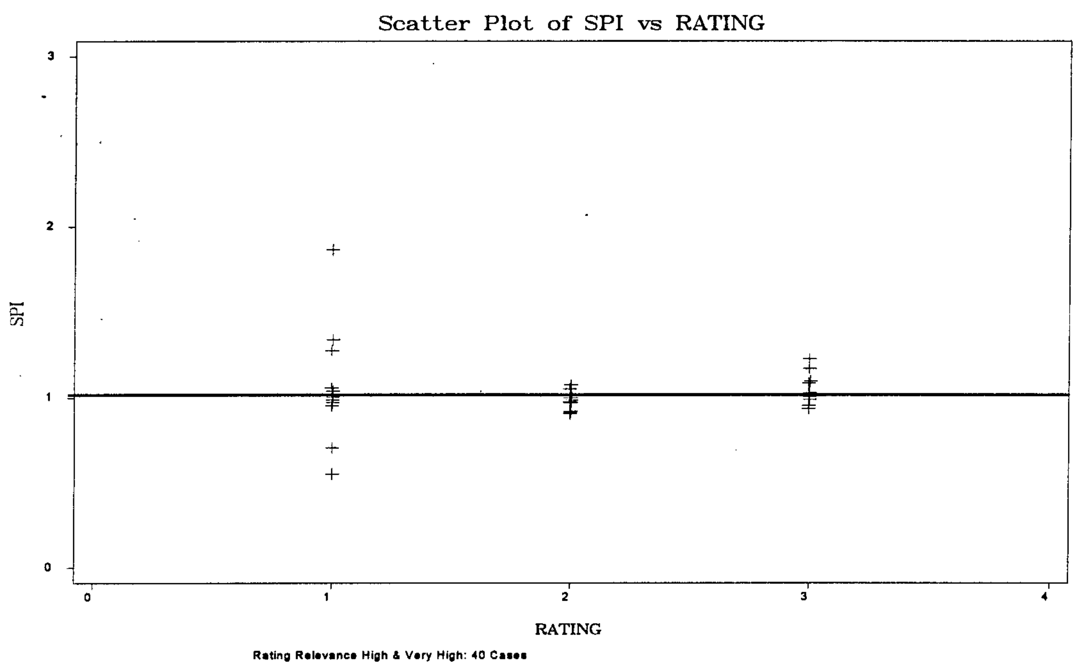

Figure D-2 Scatter Plot of SPI versus Rating for High and Very High Rating Relevance

2. Box and Whisker Plots of CPI and SPI

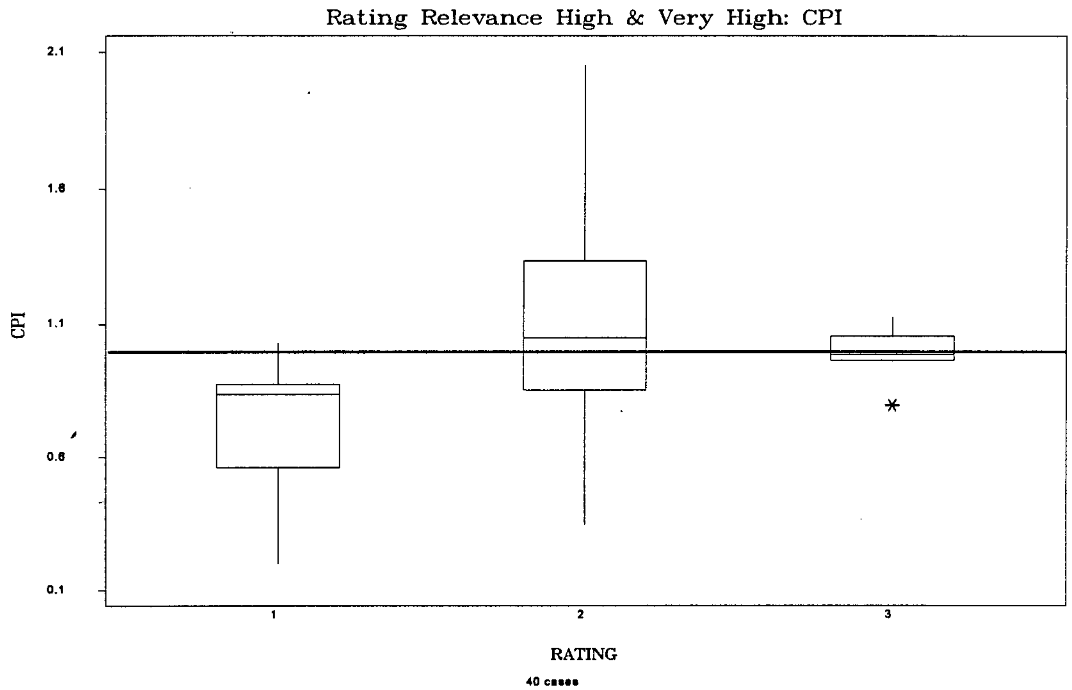

Figure D-3 Box and Whisker Plot of CPI versus Rating for High and Very High Rating Relevance

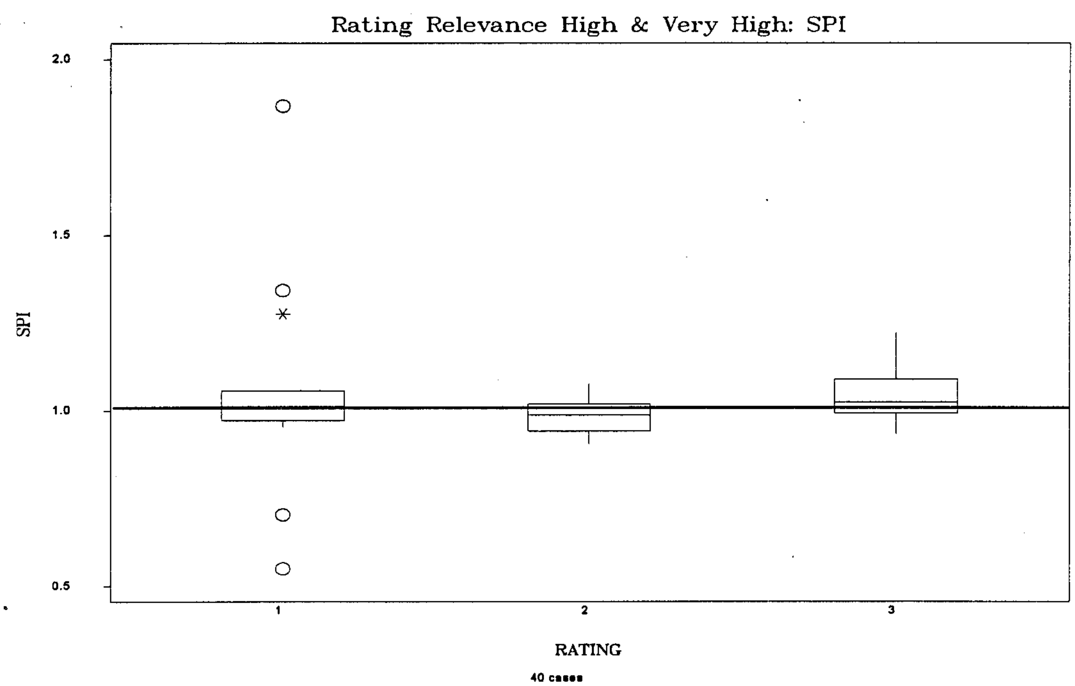

Figure D-4 Box and Whisker Plot of SPI versus Rating for High and Very High Rating Relevance

3. *Kruskal-Wallis and Multiple Comparison Tests*

Table D-1
Kruskal-Wallis for CPI for High and Very High Rating Relevance

```
KRUSKAL-WALLIS ONE-WAY NONPARAMETRIC AOV FOR CPI BY RATING

          MEAN    SAMPLE
 RATING   RANK    SIZE
   1      13.0     14
   2      24.7     15
   3      24.4     11
TOTAL     20.5     40

KRUSKAL-WALLIS STATISTIC                 8.8692
P-VALUE, USING CHI-SQUARED APPROXIMATION     0.0119
```

Table D-2
Multiple Comparison Matrix for CPI for High and Very High Rating Relevance

			Rating		
Rating	n	Mean Rank	1	2	3
1	14	13.0		--	--
2	15	24.7	3.733		--
3	11	24.4	2.761	-8.211	

K-W Statistic of 8.8692, P=0.0119

Table D-3
Kruskal-Wallis for SPI for High and Very High Rating Relevance

```
KRUSKAL-WALLIS ONE-WAY NONPARAMETRIC AOV FOR SPI BY RATING

          MEAN    SAMPLE
 RATING   RANK    SIZE
   1      21.5     14
   2      16.8     15
   3      24.3     11
TOTAL     20.5     40

KRUSKAL-WALLIS STATISTIC                 2.7738
P-VALUE, USING CHI-SQUARED APPROXIMATION     0.2498
```

Table D-4
Multiple Comparison Matrix for SPI for High and Very High Rating Relevance

			Rating		
Rating	n	Mean Rank	1	2	3
1	14	21.5		--	--
2	15	16.8	-3.267		--
3	11	24.3	-5.839	-1.011	

K-W Statistic of 2.7738, P=0.2498

4. Descriptive Statistics

Table D-5

Descriptive Statistics for High and Very High Rating Relevance

	Rating=1 CPI	Rating=1 SPI	Rating=2 CPI	Rating=2 SPI	Rating=3 CPI	Rating=3 SPI
N	14	14	15	15	11	11
Mean	0.7381	1.0558	1.1040	0.9853	0.9880	1.0457
Std Dev	0.2618	0.3039	0.4870	0.0568	0.1105	0.0891
Min	0.2019	.5507	0.3496	0.9043	0.7924	0.9335
Median	0.8357	1.0112	1.0493	0.9878	0.9856	1.0242
Max	1.0288	1.8676	2.0506	1.0774	1.1272	1.2207
MAD	0.1029	0.0417	0.2125	0.0321	0.0643	0.0622
Skew	-1.0541	1.1223	0.3738	0.0396	-0.7217	0.7397

5. Multiple Comparison Calculations for High and Very High Rating Relevance

The Kruskal-Wallis H Test Statistic

$$H := \frac{12}{N\cdot(N+1)} \cdot \sum_{j=1}^{k} \frac{\left[R_j - \frac{n_j\cdot(N+1)}{2}\right]^2}{n_j}$$

Multiple Comparison Inequality

$$\left|\overline{R_i} - \overline{R_j}\right| \le z\cdot\sqrt{\frac{N\cdot(N+1)}{12}\cdot\left(\frac{1}{n_i}+\frac{1}{n_j}\right)}$$

where

k = number of samples

R_i, R_j = sum of the ranks in the ith and jth sample, respectively

n_i, n_j = number of observations in the ith and jth sample, respectively

N = total number of observations

Mean Rank of Ratings 1, 2, & 3 for CPI

$R_{cpi} := (13.0\ \ 24.7\ \ 24.4)$

Mean Rank of Ratings 1, 2, & 3 for SPI

$R_{spi} := (21.5\ \ 16.8\ \ 24.3)$

Number of Observations in Ratings 1, 2, & 3

$n := (14\ \ 15\ \ 11)$

Calculation of N $\quad N := n_{1,1} + n_{1,2} + n_{1,3} \qquad N = 40$

Critical z Value - Corresponds to a .2 level of significance and provides at least a 80% level of confidence.
$z := 1.834$

Multiple Comparisons - "which populations differ from which others"
- If the difference in mean rank (DMR) between 2 samples is greater than the right-hand side (RHS) of the multiple comparison inequality above, there is a statistically significant difference between the 2 samples.
- The direction of the significant difference is determined by noting for each pair which sample has the larger sample mean.

Comparision of Rating 1 and 2

$DMR_{cpi} := \left|R_{cpi_{1,1}} - R_{cpi_{1,2}}\right| \qquad DMR_{cpi} = 11.7$

$DMR_{spi} := \left|R_{spi_{1,1}} - R_{spi_{1,2}}\right| \qquad DMR_{spi} = 4.7$

$RHS := z\cdot\sqrt{\frac{N\cdot(N+1)}{12}\cdot\left(\frac{1}{n_{1,1}}+\frac{1}{n_{1,2}}\right)} \qquad RHS = 7.967$

For CPI $\ DMR_{cpi} - RHS = 3.733$ — CPI at Rating 2 > CPI at Rating 1 <==

For SPI $\ DMR_{spi} - RHS = -3.267$ — no significant difference <==

Comparision of Rating 2 and 3

$DMR_{cpi} := \left|R_{cpi_{1,2}} - R_{cpi_{1,3}}\right| \qquad DMR_{cpi} = 0.3$

$DMR_{spi} := \left|R_{spi_{1,2}} - R_{spi_{1,3}}\right| \qquad DMR_{spi} = 7.5$

$RHS := z\cdot\sqrt{\frac{N\cdot(N+1)}{12}\cdot\left(\frac{1}{n_{1,2}}+\frac{1}{n_{1,3}}\right)} \qquad RHS = 8.511$

For CPI $\ DMR_{cpi} - RHS = -8.211$ — no significant difference <==

For SPI $\ DMR_{spi} - RHS = -1.011$ — no significant difference <==

Comparision of Rating 1 and 3

$DMR_{cpi} := \left|R_{cpi_{1,1}} - R_{cpi_{1,3}}\right| \qquad DMR_{cpi} = 11.4$

$DMR_{spi} := \left|R_{spi_{1,1}} - R_{spi_{1,3}}\right| \qquad DMR_{spi} = 2.8$

$RHS := z\cdot\sqrt{\frac{N\cdot(N+1)}{12}\cdot\left(\frac{1}{n_{1,1}}+\frac{1}{n_{1,3}}\right)} \qquad RHS = 8.639$

For CPI $\ DMR_{cpi} - RHS = 2.761$ — CPI at Rating 3 > CPI at Rating 1 <==

For SPI $\ DMR_{spi} - RHS = -5.839$ — no significant difference <==

Figure D-5 Calculations for Multiple Comparison Test for High and Very High Rating Relevance

Appendix E: Data Supporting the Analysis of Moderator "Rating Type"

This appendix contains the complete set of plots, tables, and calculations supporting sections 5.2.1. The plots and the Kruskal-Wallis tables for nonparametric analysis of variance are from the Statistix 4.0 computer program. The multiple comparison calculations were performed using Mathcad 4.0. Note that we did not abridge the data as we transcribed it from these computer programs into the report, and thus the number of digits reported in each calculation are not necessarily significant.

The moderator "Rating Type" is of interest because of the acknowledged difference in the results of the two rating methods, SPA and SCE (Bessleman, Byrnes, Lin, Paulk and Puranik, 1993:24). The SPA, which is primarily used for self-assessment, comprises the bulk of the data we collected. The SCE, which is performed by the government in the context of a source selection comprises only 18 of our 52 total data points.

1. Scatter Plots of CPI and SPI

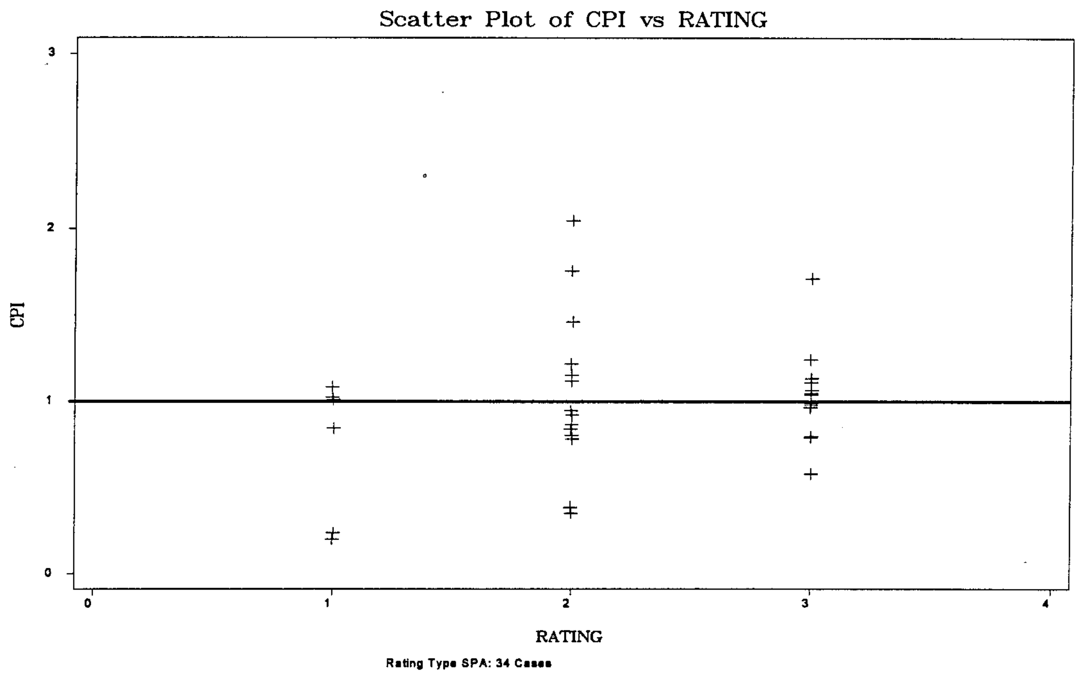

Figure E-1 Scatter Plot of CPI versus Rating for Software Process Assessment (SPA)

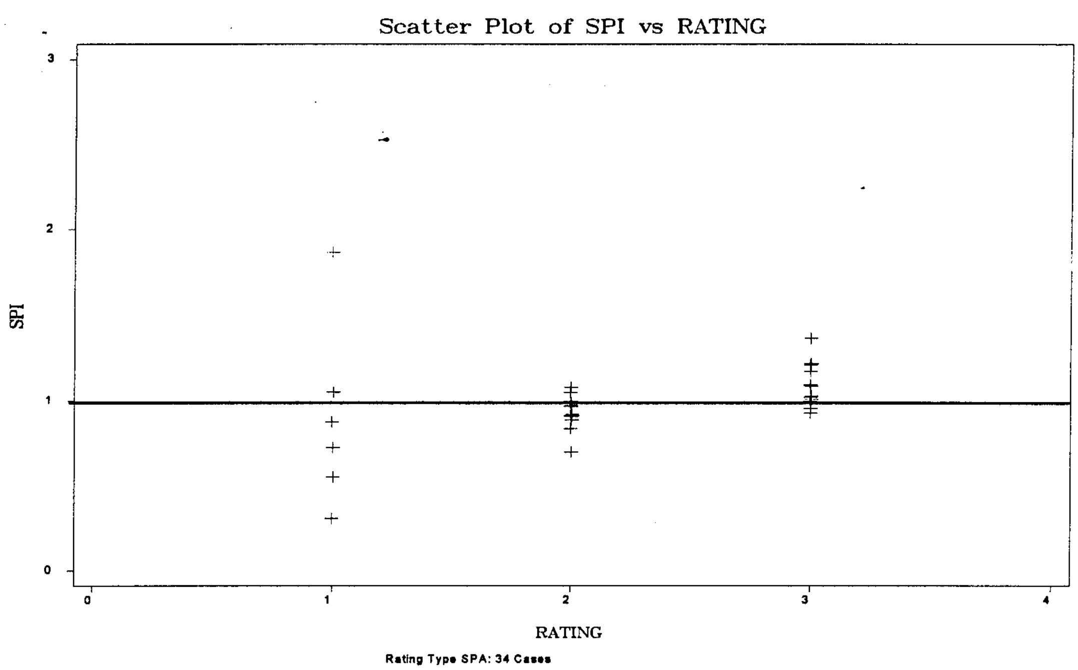

Figure E-2 Scatter Plot of SPI versus Rating for Software Process Assessment (SPA)

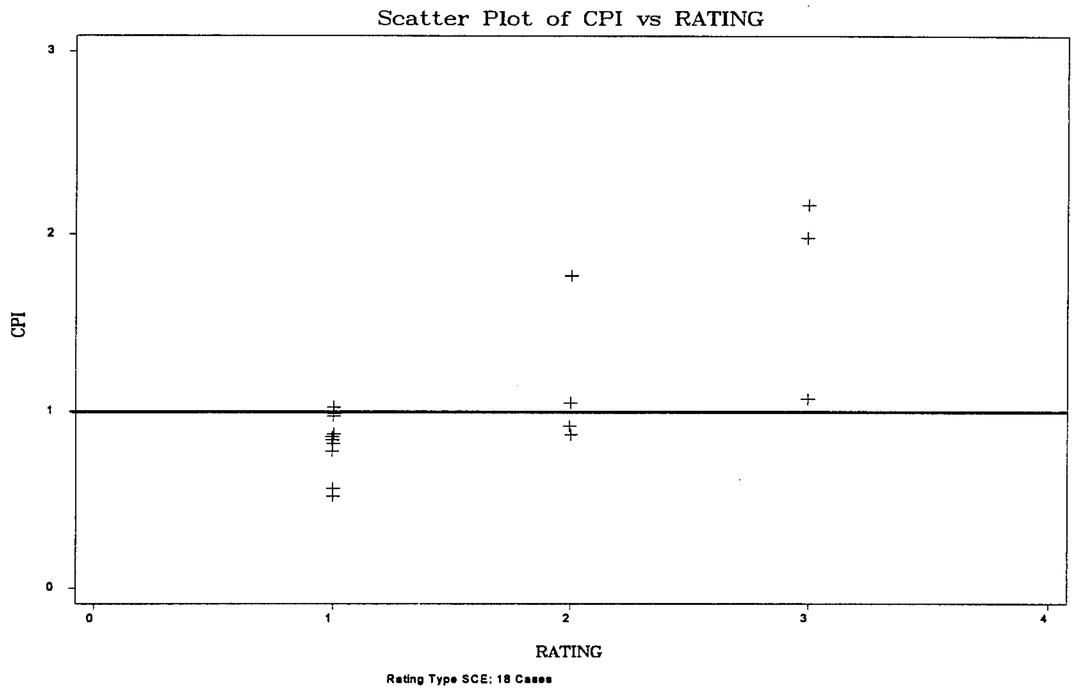

Figure E-3 Scatter Plot of CPI versus Rating for Software Capability Evaluation (SCE)

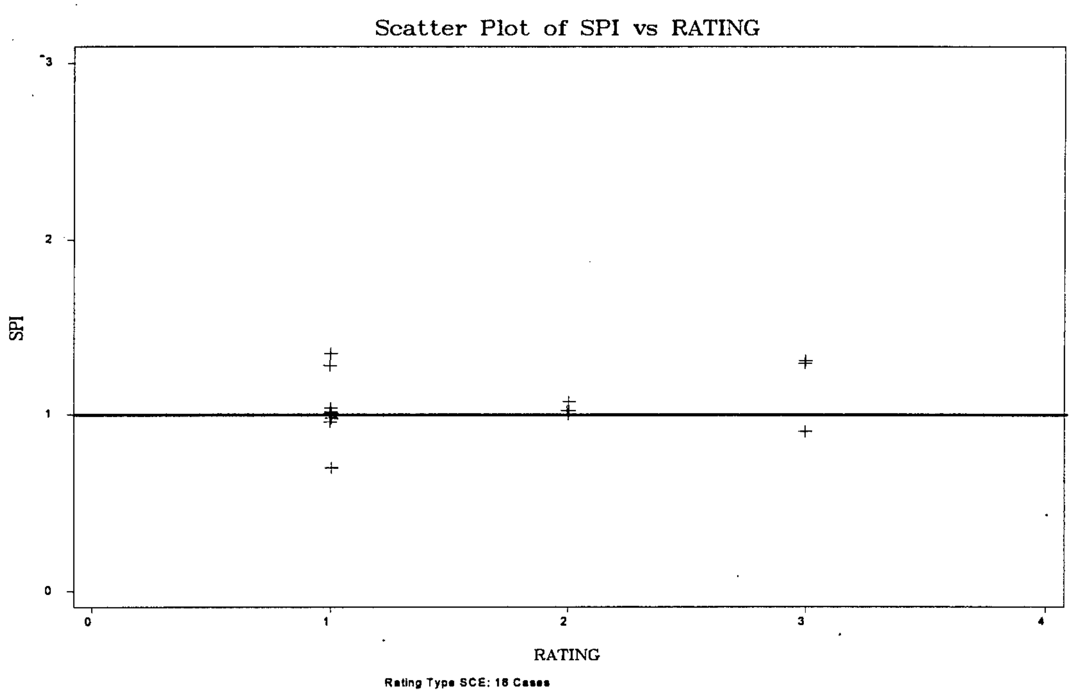

Figure E-4 Scatter Plot of SPI versus Rating for Software Capability Evaluation (SCE)

2. *Box and Whisker Plots of CPI and SPI*

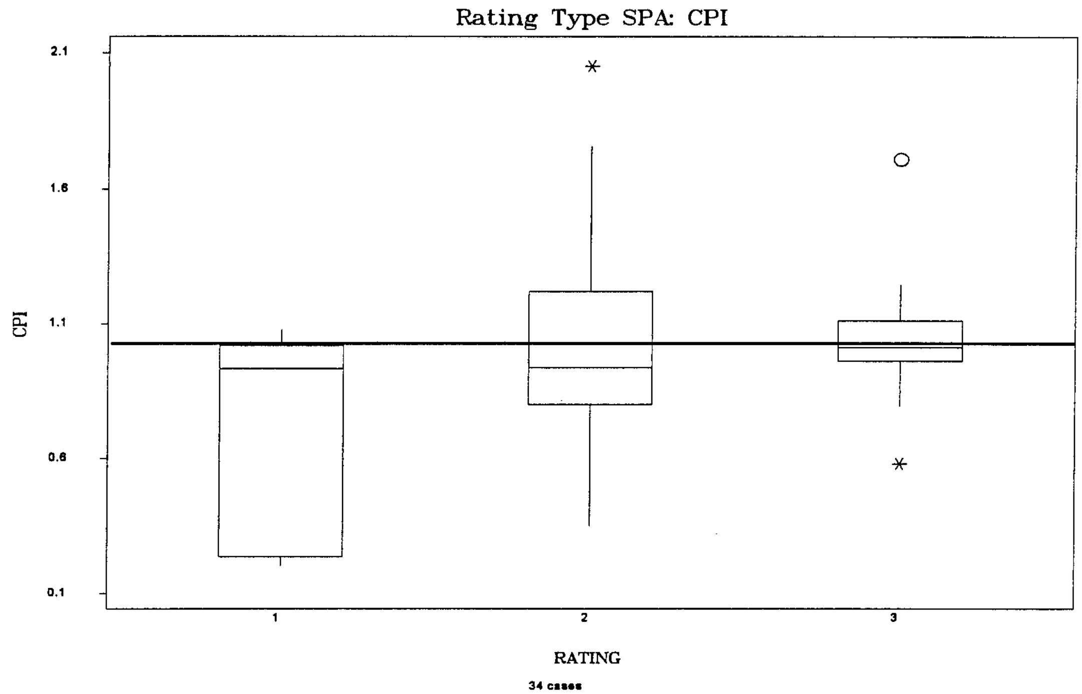

Figure E-5 Box and Whiskers Plot for CPI for Software Process Assessment (SPA)

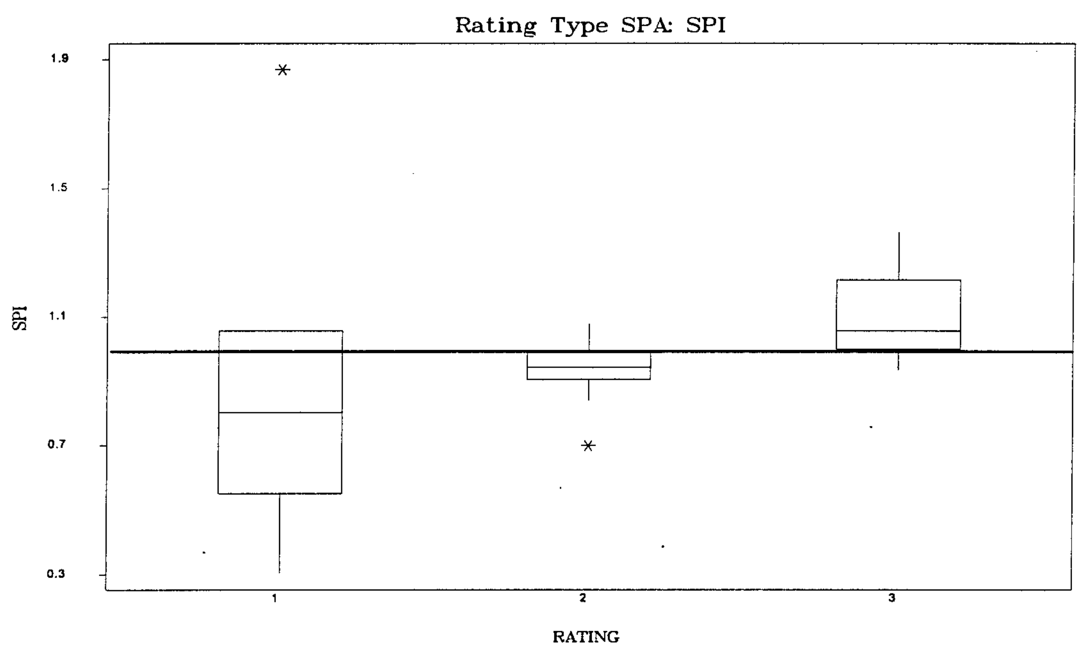

Figure E-6 Box and Whiskers Plot for SPI for Software Process Assessment (SPA)

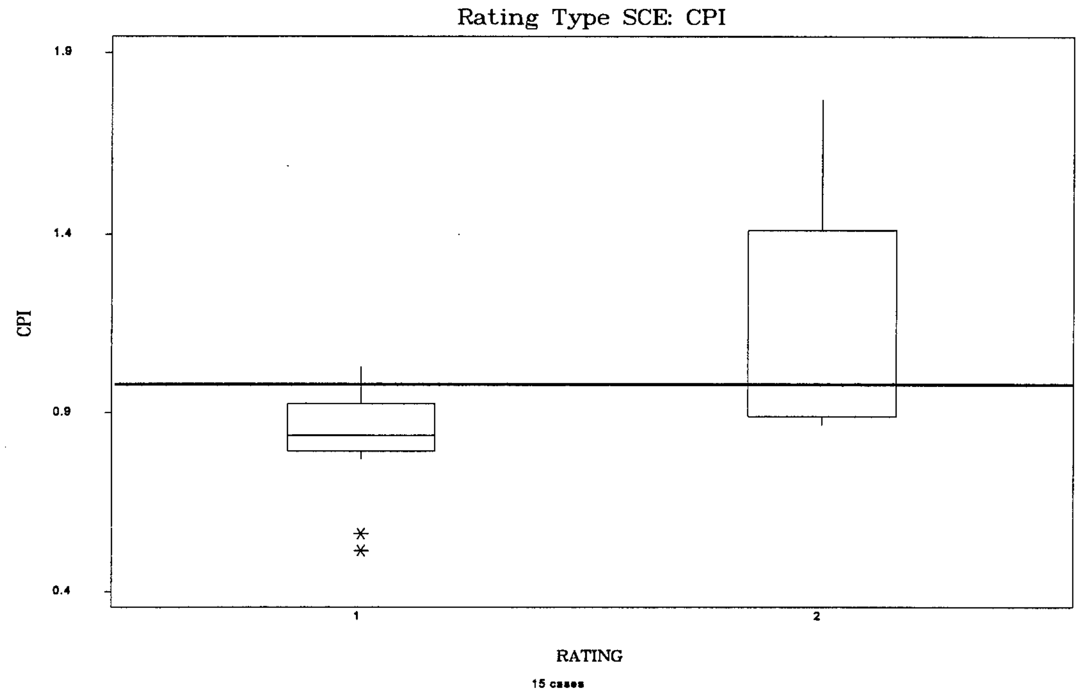

Figure E-7 Box and Whiskers Plot for CPI for Software Capability Evaluation (SCE)

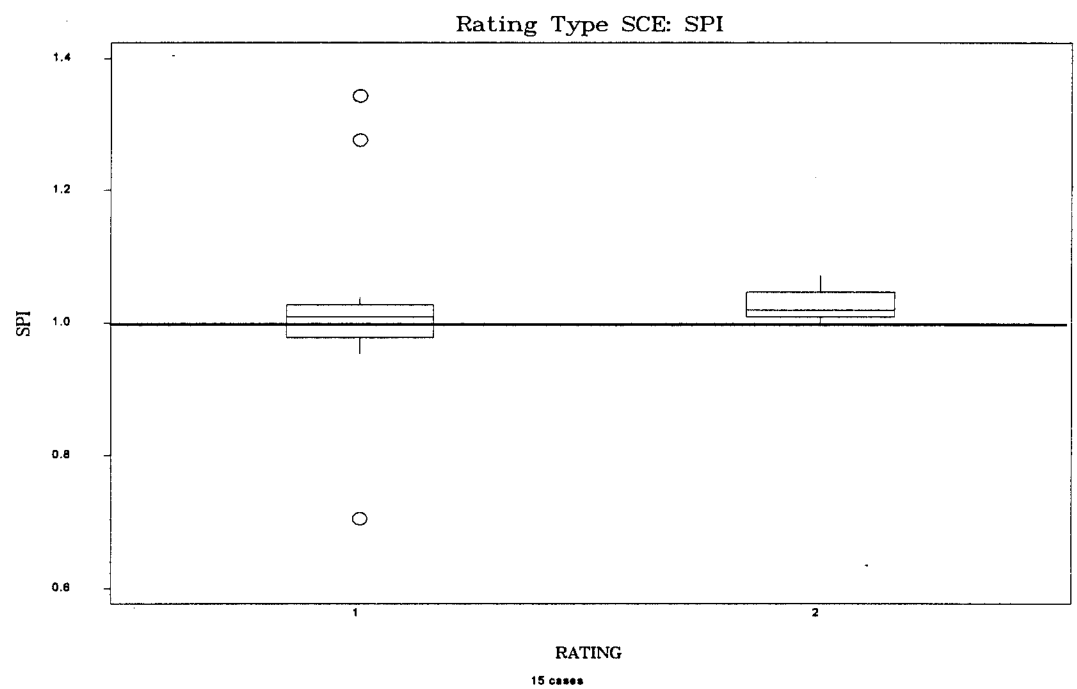

Figure E-8 Box and Whiskers Plot for SPI for Software Capability Evaluation (SCE)

3. *Kruskal-Wallis and Multiple Comparison Tests*

Table E-1
Kruskal-Wallis for CPI for Software Process Assessment (SPA)

KRUSKAL-WALLIS ONE-WAY NONPARAMETRIC AOV FOR CPI BY RATING

RATING	MEAN RANK	SAMPLE SIZE
1	12.8	6
2	18.0	14
3	19.0	14
TOTAL	17.5	34

KRUSKAL-WALLIS STATISTIC 1.6706
P-VALUE, USING CHI-SQUARED APPROXIMATION 0.4337

Table E-2
Multiple Comparison Matrix for CPI for Software Process Assessment (SPA)

			Rating		
Rating	n	Mean Rank	1	2	3
1	6	12.8		--	--
2	14	18.0	-3.712		--
3	14	19.0	-2.712	-5.903	

K-W Statistic of 1.6706, P=0.4337

Table E-3
Kruskal-Wallis for SPI for Software Process Assessment (SPA)

KRUSKAL-WALLIS ONE-WAY NONPARAMETRIC AOV FOR SPI BY RATING

RATING	MEAN RANK	SAMPLE SIZE
1	12.0	6
2	13.3	14
3	24.1	14
TOTAL	17.5	34

KRUSKAL-WALLIS STATISTIC 10.5448
P-VALUE, USING CHI-SQUARED APPROXIMATION 0.0051

Table E-4
Multiple Comparison Matrix for SPI for Software Process Assessment (SPA)

			Rating		
Rating	n	Mean Rank	1	2	3
1	6	12.0		--	--
2	14	13.3	-7.612		--
3	14	24.1	3.188	3.897	

K-W Statistic of 10.5448, P=0.0051

Table E-5
Kruskal-Wallis for CPI for Software Capability Evaluation (SCE)

```
KRUSKAL-WALLIS ONE-WAY NONPARAMETRIC AOV FOR CPI BY RATING

         MEAN   SAMPLE
RATING   RANK   SIZE
     1    6.6     11
     2   12.0      4
     3   16.7      3
TOTAL     9.5     18

KRUSKAL-WALLIS STATISTIC              9.4487
P-VALUE, USING CHI-SQUARED APPROXIMATION    0.0089
```

Table E-6
Multiple Comparison Matrix for CPI for Software Capability Evaluation (SCE)

			Rating		
Rating	n	Mean Rank	1	2	3
1	11	6.6		--	--
2	4	12.0	-0.3173		--
3	3	16.7	3.723	-2.778	

K-W Statistic of 9.4487, P=0.0089

Table E-7
Kruskal-Wallis for SPI for Software Capability Evaluation (SCE)

```
KRUSKAL-WALLIS ONE-WAY NONPARAMETRIC AOV FOR SPI BY RATING

         MEAN   SAMPLE
RATING   RANK   SIZE
     1    8.4     11
     2   10.9      4
     3   11.7      3
TOTAL     9.5     18

KRUSKAL-WALLIS STATISTIC              1.2201
P-VALUE, USING CHI-SQUARED APPROXIMATION    0.5433
```

Table E-8
Multiple Comparison Matrix for SPI for Software Capability Evaluation (SCE)

			Rating		
Rating	n	Mean Rank	1	2	3
1	11	8.4		--	--
2	4	10.9	-3.217		--
3	3	11.7	-3.077	-6.678	

K-W Statistic of 1.2201, P=0.5433

4. Descriptive Statistics

Table E-9
Descriptive Statistics for Software Process Assessment (SPA)

	Rating=1 CPI	Rating=1 SPI	Rating=2 CPI	Rating=2 SPI	Rating=3 CPI	Rating=3 SPI
N	6	6	14	14	14	14
Mean	0.7333	0.8967	1.0451	0.9359	1.0288	1.0929
Std Dev	0.4057	0.5420	0.4703	0.0933	0.2559	.01267
Min	0.2019	0.3028	0.3496	0.6978	0.5808	0.9335
Median	0.9321	0.8015	0.9365	0.9423	1.0104	1.0563
Max	1.0788	1.8676	2.0506	1.0774	1.7097	1.3652
MAD	0.1166	0.2527	0.1931	0.0418	0.0754	0.0860
Skew	-0.6058	0.9136	0.5967	-0.9626	1.0047	0.6501

Table E-10
Descriptive Statistics for Software Capability Evaluation (SCE)

	Rating=1 CPI	Rating=1 SPI	Rating=2 CPI	Rating=2 SPI	Rating=3 CPI	Rating=3 SPI
N	11	11	4	4	3	3
Mean	0.8223	1.0280	1.1505	1.0275	1.7365	1.1664
Std Dev	0.1615	0.1666	0.4227	0.0309	0.5862	0.2310
Min	0.5143	0.7044	0.8666	1.0000	1.0675	0.8998
Median	0.8380	1.0090	0.9808	1.01090	1.9819	1.2920
Max	1.0228	1.3428	1.7738	1.0718	2.1602	1.3073
MAD	0.0669	0.0294	0.0913	0.0010	.01783	0.0153
Skew	-0.7512	0.2689	1.0432	0.8506	-0.6343	-0.7036

5. Multiple Comparison Calculations for Rating Type - SPA

The Kruskal-Wallis H Test Statistic

$$H := \frac{12}{N \cdot (N+1)} \cdot \sum_{j=1}^{k} \frac{\left[R_j - \frac{n_j \cdot (N+1)}{2}\right]^2}{n_j}$$

Multiple Comparison Inequality

$$\left|\overline{R_i} - \overline{R_j}\right| \le z \cdot \sqrt{\frac{N \cdot (N+1)}{12} \cdot \left(\frac{1}{n_i} + \frac{1}{n_j}\right)}$$

where

k = number of samples

R_i, R_j = sum of the ranks in the ith and jth sample, respectively

n_i, n_j = number of observations in the ith and jth sample, respectively

N = total number of observations

Mean Rank of Ratings 1, 2, & 3 for CPI

$R_{cpi} := (12.8 \;\; 18.0 \;\; 19.0)$

Mean Rank of Ratings 1, 2, & 3 for SPI

$R_{spi} := (12.0 \;\; 13.3 \;\; 24.1)$

Number of Observations in Ratings 1, 2, & 3

$n := (6 \;\; 14 \;\; 14)$

Calculation of N $\quad N := n_{1,1} + n_{1,2} + n_{1,3} \qquad N = 34$

Critical z Value - Corresponds to a .2 level of significance and provides at least a 80% level of confidence.

$z := 1.834$

Multiple Comparisons - "which populations differ from which others"

- If the difference in mean rank (DMR) between 2 samples is greater than the right-hand side (RHS) of the multiple comparison inequality above, there is a statistically significant difference between the 2 samples.
- The direction of the significant difference is determined by noting for each pair which sample has the larger sample mean.

Comparision of Rating 1 and 2

$DMR_{cpi} := \left|R_{cpi_{1,1}} - R_{cpi_{1,2}}\right| \qquad DMR_{cpi} = 5.2$

$DMR_{spi} := \left|R_{spi_{1,1}} - R_{spi_{1,2}}\right| \qquad DMR_{spi} = 1.3$

$RHS := z \cdot \sqrt{\frac{N \cdot (N+1)}{12} \cdot \left(\frac{1}{n_{1,1}} + \frac{1}{n_{1,2}}\right)} \qquad RHS = 8.912$

For CPI $DMR_{cpi} - RHS = -3.712$ — no significant difference <==

For SPI $DMR_{spi} - RHS = -7.612$ — no significant difference <==

Comparision of Rating 2 and 3

$DMR_{cpi} := \left|R_{cpi_{1,2}} - R_{cpi_{1,3}}\right| \qquad DMR_{cpi} = 1$

$DMR_{spi} := \left|R_{spi_{1,2}} - R_{spi_{1,3}}\right| \qquad DMR_{spi} = 10.8$

$RHS := z \cdot \sqrt{\frac{N \cdot (N+1)}{12} \cdot \left(\frac{1}{n_{1,2}} + \frac{1}{n_{1,3}}\right)} \qquad RHS = 6.903$

For CPI $DMR_{cpi} - RHS = -5.903$ — no significant difference <==

For SPI $DMR_{spi} - RHS = 3.897$ — SPI at Rating 3 > SPI at Rating 2 <==

Comparision of Rating 1 and 3

$DMR_{cpi} := \left|R_{cpi_{1,1}} - R_{cpi_{1,3}}\right| \qquad DMR_{cpi} = 6.2$

$DMR_{spi} := \left|R_{spi_{1,1}} - R_{spi_{1,3}}\right| \qquad DMR_{spi} = 12.1$

$RHS := z \cdot \sqrt{\frac{N \cdot (N+1)}{12} \cdot \left(\frac{1}{n_{1,1}} + \frac{1}{n_{1,3}}\right)} \qquad RHS = 8.912$

For CPI $DMR_{cpi} - RHS = -2.712$ — no significant difference <==

For SPI $DMR_{spi} - RHS = 3.188$ — SPI at Rating 3 > SPI at Rating 1 <==

Figure E-9 Calculations for Multiple Comparison Test for Software Process Assessment (SPA)

6. *Multiple Comparison Calculations for Rating Type - SCE*

The Kruskal-Wallis H Test Statistic

$$H := \frac{12}{N \cdot (N+1)} \cdot \sum_{j=1}^{k} \frac{\left[R_j - \frac{n_j \cdot (N+1)}{2}\right]^2}{n_j}$$

Multiple Comparison Inequality

$$\left|\overline{R_i} - \overline{R_j}\right| \le z \cdot \sqrt{\frac{N \cdot (N+1)}{12} \cdot \left(\frac{1}{n_i} + \frac{1}{n_j}\right)}$$

where
k = number of samples
R_i, R_j = sum of the ranks in the ith and jth sample, respectively
n_i, n_j = number of observations in the ith and jth sample, respectively
N = total number of observations

Mean Rank of Ratings 1, 2, & 3 for CPI
$R_{cpi} := (6.6 \;\; 12.0 \;\; 16.7)$

Mean Rank of Ratings 1, 2, & 3 for SPI
$R_{spi} := (8.4 \;\; 10.9 \;\; 11.7)$

Number of Observations in Ratings 1, 2, & 3
$n := (11 \;\; 4 \;\; 3)$

Calculation of N $\quad N := n_{1,1} + n_{1,2} + n_{1,3} \quad N = 18$

Critical z Value - Corresponds to a .2 level of significance and provides at least a 80% level of confidence.
$z := 1.834$

<u>**Multiple Comparisons**</u> - "which populations differ from which others"
- If the difference in mean rank (DMR) between 2 samples is greater than the right-hand side (RHS) of the multiple comparison inequality above, there is a statistically significant difference between the 2 samples.
- The direction of the significant difference is determined by noting for each pair which sample has the larger sample mean.

Comparision of Rating 1 and 2

$DMR_{cpi} := \left|R_{cpi_{1,1}} - R_{cpi_{1,2}}\right| \quad DMR_{cpi} = 5.4$

$DMR_{spi} := \left|R_{spi_{1,1}} - R_{spi_{1,2}}\right| \quad DMR_{spi} = 2.5$

$RHS := z \cdot \sqrt{\frac{N \cdot (N+1)}{12} \cdot \left(\frac{1}{n_{1,1}} + \frac{1}{n_{1,2}}\right)} \quad RHS = 5.717$

For CPI $\; DMR_{cpi} - RHS = -0.317$ — no significant difference <===
For SPI $\; DMR_{spi} - RHS = -3.217$ — no significant difference <===

Comparision of Rating 2 and 3

$DMR_{cpi} := \left|R_{cpi_{1,2}} - R_{cpi_{1,3}}\right| \quad DMR_{cpi} = 4.7$

$DMR_{spi} := \left|R_{spi_{1,2}} - R_{spi_{1,3}}\right| \quad DMR_{spi} = 0.8$

$RHS := z \cdot \sqrt{\frac{N \cdot (N+1)}{12} \cdot \left(\frac{1}{n_{1,2}} + \frac{1}{n_{1,3}}\right)} \quad RHS = 7.478$

For CPI $\; DMR_{cpi} - RHS = -2.778$ — no significant difference <===
For SPI $\; DMR_{spi} - RHS = -6.678$ — no significant difference <===

Comparision of Rating 1 and 3

$DMR_{cpi} := \left|R_{cpi_{1,1}} - R_{cpi_{1,3}}\right| \quad DMR_{cpi} = 10.1$

$DMR_{spi} := \left|R_{spi_{1,1}} - R_{spi_{1,3}}\right| \quad DMR_{spi} = 3.3$

$RHS := z \cdot \sqrt{\frac{N \cdot (N+1)}{12} \cdot \left(\frac{1}{n_{1,1}} + \frac{1}{n_{1,3}}\right)} \quad RHS = 6.377$

For CPI $\; DMR_{cpi} - RHS = 3.723$ — CPI at Rating 3 > CPI at Rating 1 <===
For SPI $\; DMR_{spi} - RHS = -3.077$ — no significant difference <===

Figure E-10 Calculations for Multiple Comparison Test for Software Capability Evaluation (SCE)

Appendix F: Data Supporting the Analysis of Moderator "Baseline Volatility"

This appendix contains the complete set of plots, tables, and calculations supporting sections 5.2.2. The plots and the Kruskal-Wallis tables for nonparametric analysis of variance are from the Statistix 4.0 computer program. The multiple comparison calculations were performed using Mathcad 4.0. Note that we did not abridge the data as we transcribed it from these computer programs into the report, and thus the number of digits reported in each calculation are not necessarily significant.

The moderator "Baseline Volatility" is defined as the proportional change in the budget-at-complete (BAC) over the period of interest, i.e., we calculated the change in total budget over the 12 month period as a percentage of the budget at the beginning of the period. This rate of change of the budget is indicative of rebaselining, whatever the source, whether it is due to reallocation of work, ECPs, or reprogramming. We arbitrarily selected a change in budget of plus or minus fifteen percent as the stratification level in our analysis.

This moderator was considered because a correlation between rating and performance could be affected by the relative changes in the baseline of a project. The causes of these changes in baseline can vary from redirection on behalf of the government to inadequate initial budgeting by the contractor. It is possible that a change in the baseline, regardless of the type, could affect the link between contractor performance and the performance indices of interest.

1. Scatter Plots of CPI and SPI

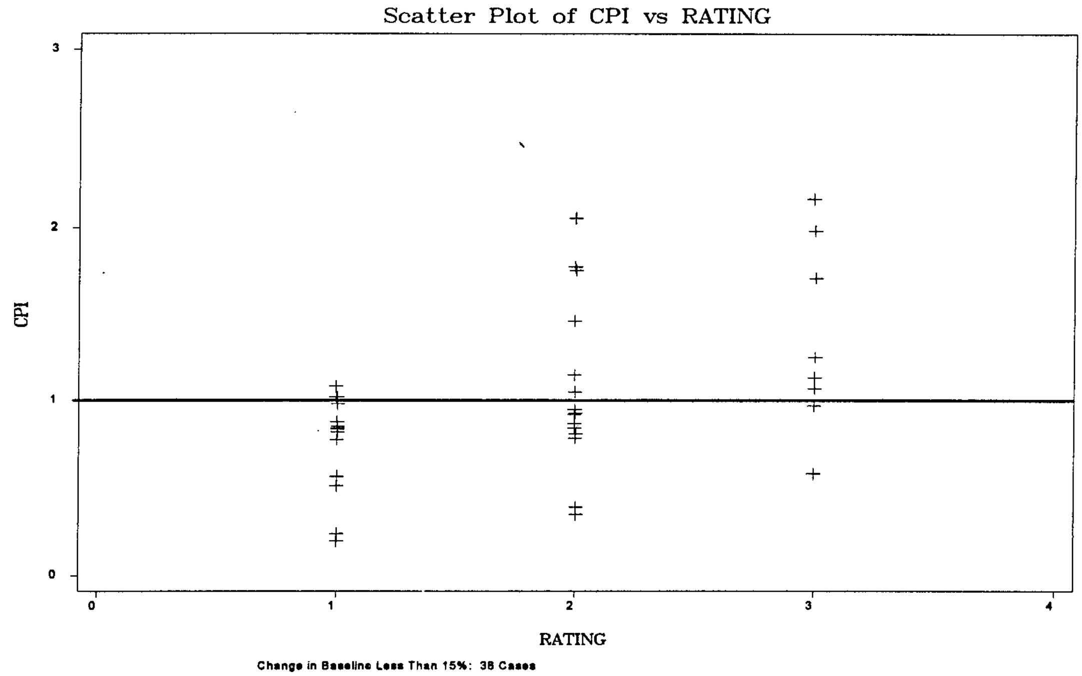

Figure F-1 Scatter Plot of CPI versus Rating for Less than 15% Change in Baseline

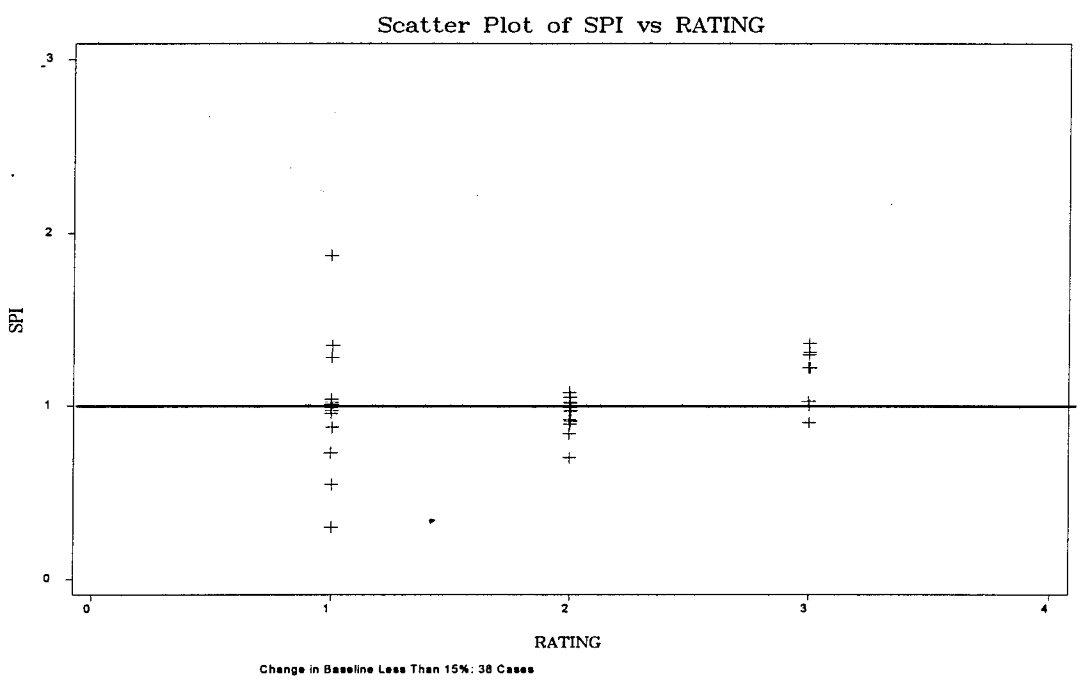

Figure F-2 Scatter Plot of SPI versus Rating for Less than 15% Change in Baseline

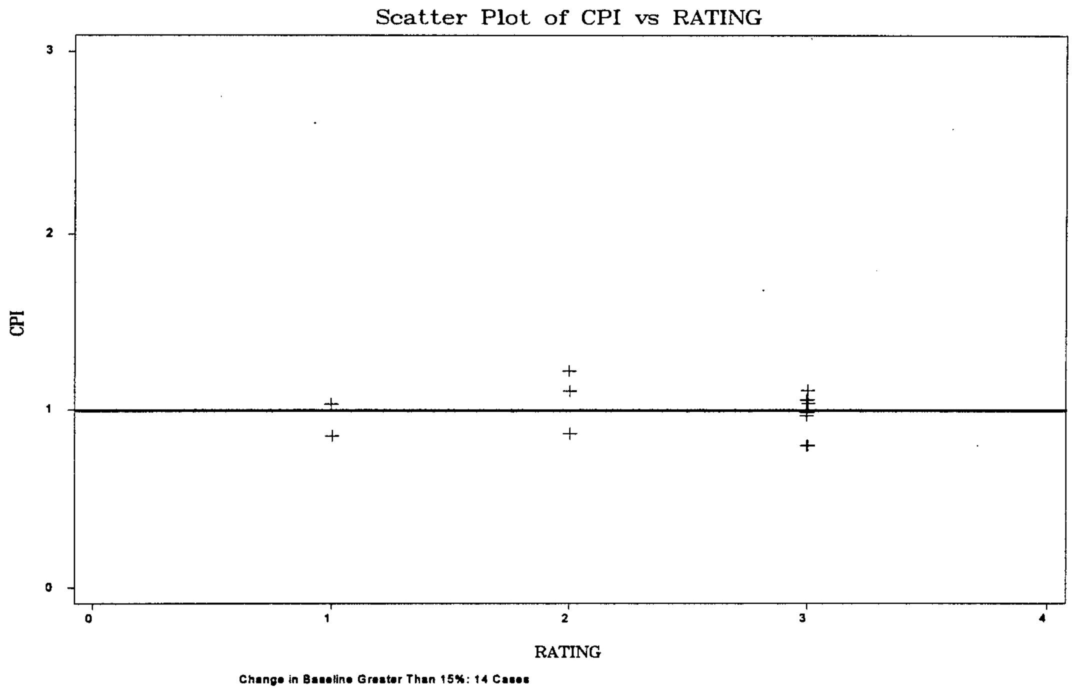

Figure F-3 Scatter Plot of CPI versus Rating for Greater than 15% Change in Baseline

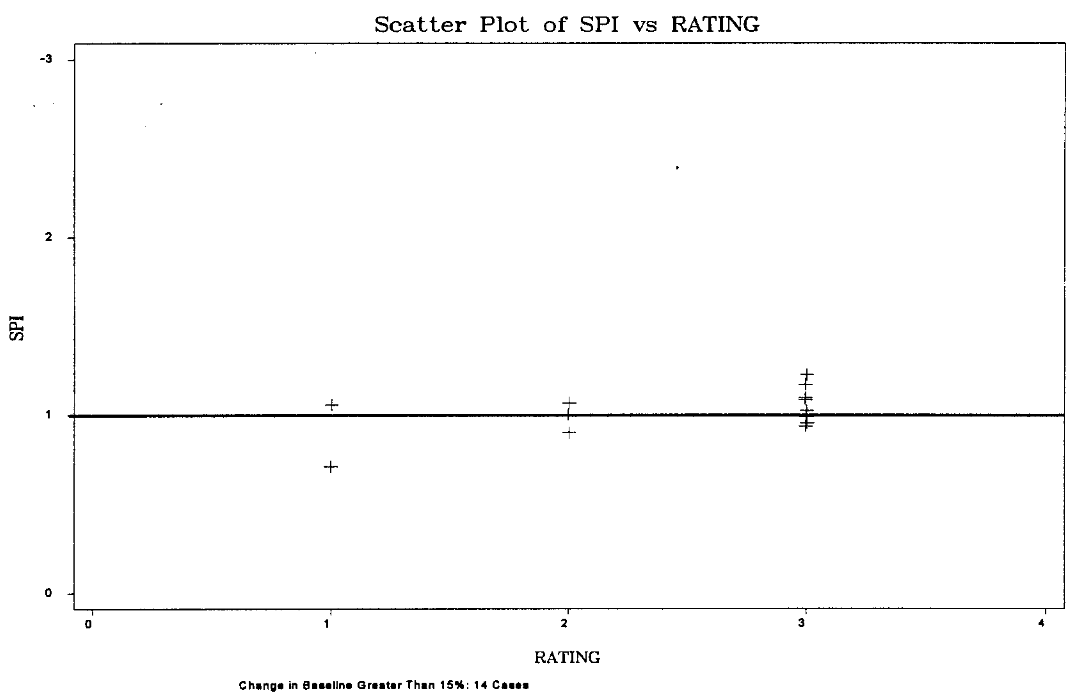

Figure F-4 Scatter Plot of SPI versus Rating for Greater than 15% Change in Baseline

2. Box and Whisker Plots of CPI and SPI

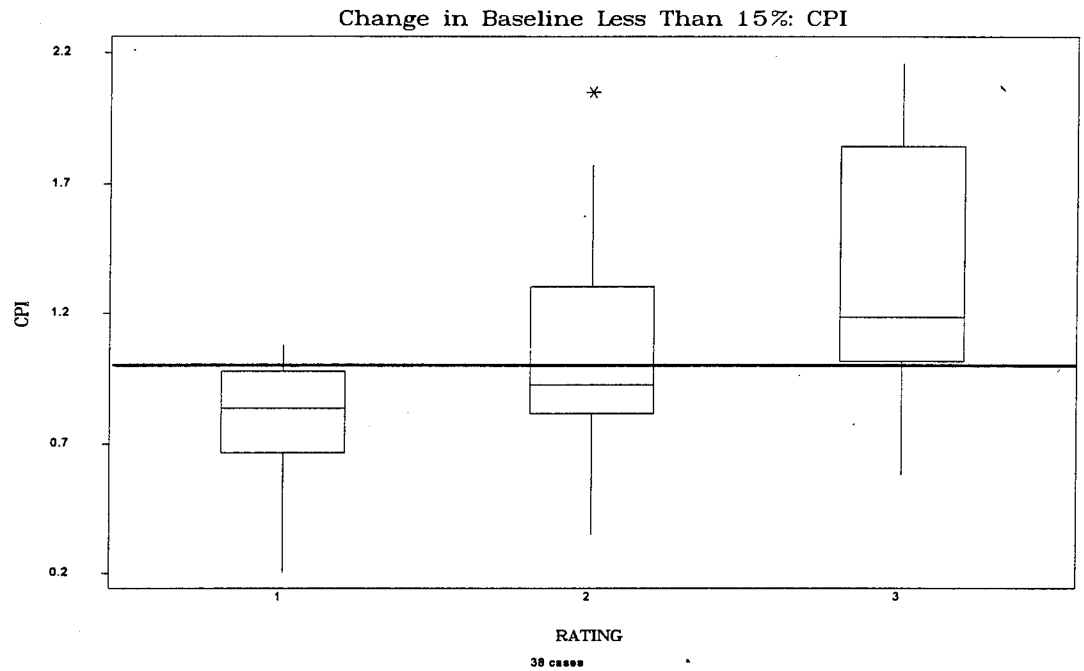

Figure F-5 Box and Whiskers for CPI for Less than 15% Change in Baseline

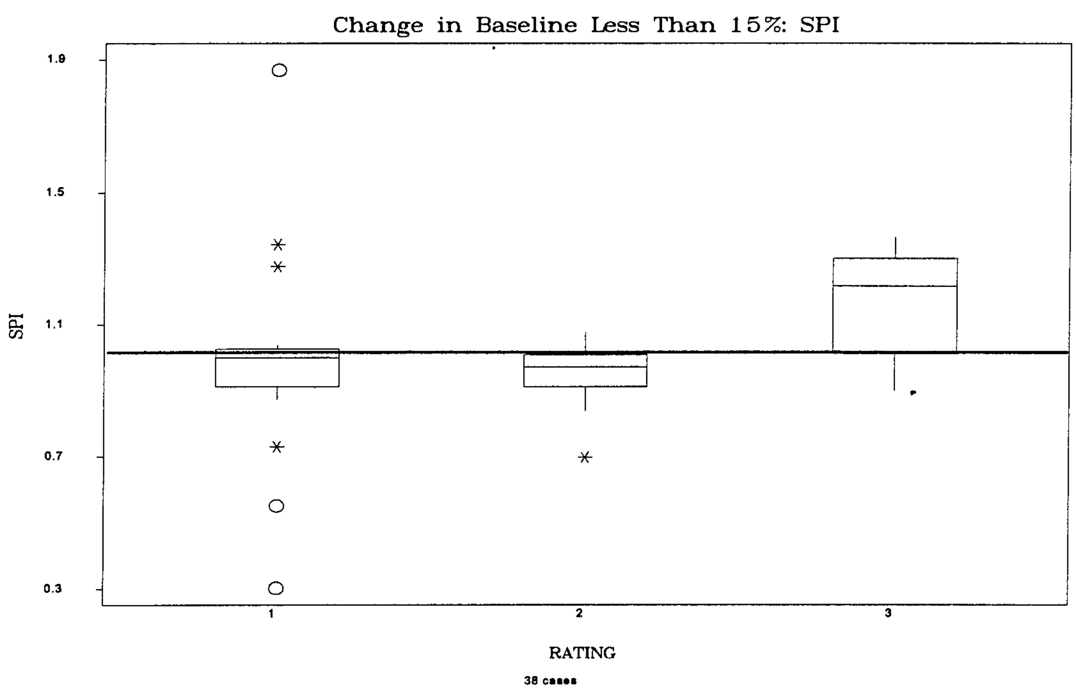

Figure F-6 Box and Whiskers for SPI for Less than 15% Change in Baseline

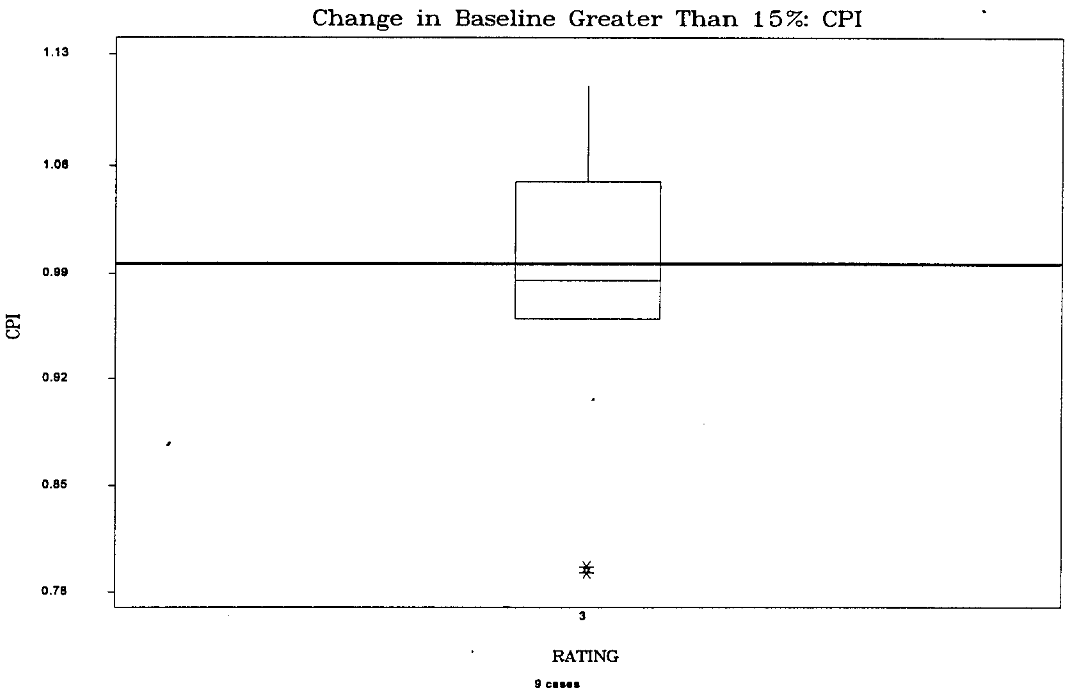

Figure F-7 Box and Whiskers for CPI for Greater than 15% Change in Baseline

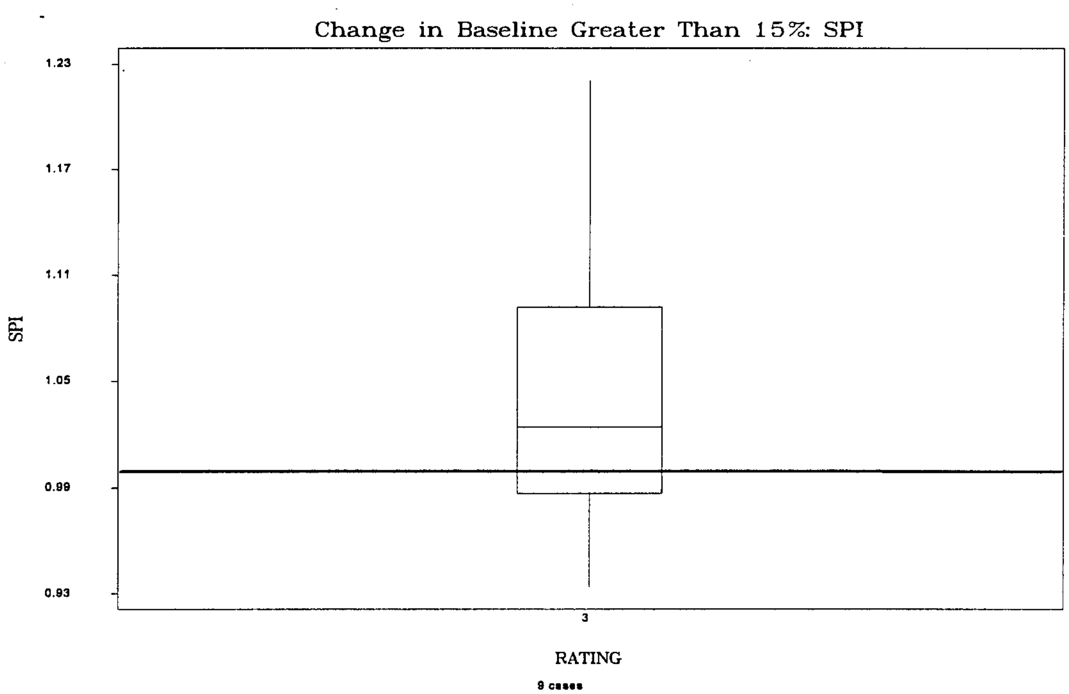

Figure F-8 Box and Whiskers for SPI for Greater than 15% Change in Baseline

3. Kruskal-Wallis and Multiple Comparison Tests

Table F-1
Kruskal-Wallis for CPI Less than 15% Change in Baseline

```
KRUSKAL-WALLIS ONE-WAY NONPARAMETRIC AOV FOR CPI BY RATING

          MEAN   SAMPLE
 RATING   RANK   SIZE
      1   14.2     15
      2   20.4     15
      3   27.8      8
TOTAL     19.5     38

KRUSKAL-WALLIS STATISTIC              7.9190
P-VALUE, USING CHI-SQUARED APPROXIMATION    0.0191
```

Table F-2
Multiple Comparison Matrix for CPI for Less than 15% Change in Baseline

			Rating		
Rating	n	Mean Rank	1	2	3
1	15	14.2		--	--
2	15	20.4	-1.242		--
3	8	27.8	4.677	-1.523	

K-W Statistic of 7.9190, P=0.0191

Table F-3
Kruskal-Wallis for SPI for Less than 15% Change in Baseline

```
KRUSKAL-WALLIS ONE-WAY NONPARAMETRIC AOV FOR SPI BY RATING

          MEAN   SAMPLE
 RATING   RANK   SIZE
      1   18.7     15
      2   15.8     15
      3   28.0      8
TOTAL     19.5     38

KRUSKAL-WALLIS STATISTIC              6.4301
P-VALUE, USING CHI-SQUARED APPROXIMATION    0.0402
```

Table F-4
Multiple Comparison Matrix for SPI for Less than 15% Change in Baseline

			Rating		
Rating	n	Mean Rank	1	2	3
1	15	18.7		--	--
2	15	15.8	-4.542		--
3	8	28.0	0.377	3.277	

K-W Statistic of 6.4301, P=0.0402

Table F-5
Kruskal-Wallis for CPI for Greater than 15% Change in Baseline

KRUSKAL-WALLIS ONE-WAY NONPARAMETRIC AOV FOR CPI BY RATING

RATING	MEAN RANK	SAMPLE SIZE
1	5.5	2
2	10.3	3
3	7.0	9
TOTAL	7.5	14

KRUSKAL-WALLIS STATISTIC 1.9619
P-VALUE, USING CHI-SQUARED APPROXIMATION 0.3750

Table F-6
Multiple Comparison Matrix for CPI for Greater than 15% Change in Baseline

			Rating		
Rating	n	Mean Rank	1	2	3
1	2	5.5		--	--
2	3	10.3	-2.204		--
3	9	7.0	-4.498	-1.815	

K-W Statistic of 1.9619, P=0.3750

Table F-7
Kruskal-Wallis for SPI for Greater than 15% Change in Baseline

KRUSKAL-WALLIS ONE-WAY NONPARAMETRIC AOV FOR SPI BY RATING

RATING	MEAN RANK	SAMPLE SIZE
1	5.0	2
2	6.0	3
3	8.6	9
TOTAL	7.5	14

KRUSKAL-WALLIS STATISTIC 1.6730
P-VALUE, USING CHI-SQUARED APPROXIMATION 0.4332

Table F-8
Multiple Comparison Matrix for SPI for Greater than 15% Change in Baseline

			Rating		
Rating	n	Mean Rank	1	2	3
1	2	5.0		--	--
2	3	6.0	-6.004		--
3	9	8.6	-2.398	-2.515	

K-W Statistic of 1.6730, P=0.4332

4. Descriptive Statistics

Table F-9
Descriptive Statistics for Less than 15% Change in Baseline

	Rating=1 CPI	Rating=1 SPI	Rating=2 CPI	Rating=2 SPI	Rating=3 CPI	Rating=3 SPI
N	15	15	15	15	8	8
Mean	0.7711	0.9951	1.0691	0.9491	1.3549	1.1654
Std Dev	0.2737	0.3513	0.4914	0.0939	0.5429	0.1684
Min	0.2019	0.3028	0.3496	0.6978	0.5808	0.8998
Median	0.8380	1.0000	0.9273	0.9726	1.1859	1.2164
Max	1.0788	1.8676	2.0506	1.0774	2.1602	1.3652
MAD	0.1439	0.0465	0.1481	0.0541	0.3712	0.1199
Skew	-1.0260	0.4913	0.5684	-1.2090	0.2427	-0.4102

Table F-10
Descriptive Statistics for Greater than 15% Change in Baseline

	Rating=1 CPI	Rating=1 SPI	Rating=2 CPI	Rating=2 SPI	Rating=3 CPI	Rating=3 SPI
N	2	2	3	3	9	9
Mean	0.9393	0.8802	1.0653	0.9920	0.9748	1.0530
Std Dev	0.1265	0.2486	0.1801	0.0840	0.1122	0.0978
Min	0.8498	0.7044	0.8666	0.9043	0.7924	0.9335
Median	0.9393	0.8802	1.1114	1.0000	0.9856	1.0242
Max	1.0288	1.0560	1.2180	1.0718	1.1105	1.2207
MAD	0.0895	0.1758	0.1066	0.0718	0.0643	0.0677
Skew	0.00	0.00	-0.4389	-0.1727	-0.7484	0.4861

5. Multiple Comparison Calculations for Less Than 15% Change in Baseline

The Kruskal-Wallis H Test Statistic

$$H := \frac{12}{N\cdot(N+1)}\cdot\sum_{j=1}^{k}\frac{\left[R_j - \frac{n_j\cdot(N+1)}{2}\right]^2}{n_j}$$

Multiple Comparison Inequality

$$\left|\overline{R_i} - \overline{R_j}\right| \le z\cdot\sqrt{\frac{N\cdot(N+1)}{12}\cdot\left(\frac{1}{n_i}+\frac{1}{n_j}\right)}$$

where

k = number of samples

R_i, R_j = sum of the ranks in the ith and jth sample, respectively

n_i, n_j = number of observations in the ith and jth sample, respectively

N = total number of observations

Mean Rank of Ratings 1, 2, & 3 for CPI

$R_{cpi} := (14.2\ \ 20.4\ \ 27.8)$

Mean Rank of Ratings 1, 2, & 3 for SPI

$R_{spi} := (18.7\ \ 15.8\ \ 28.0)$

Number of Observations in Ratings 1, 2, & 3

$n := (15\ \ 15\ \ 8)$

Calculation of N $\quad N := n_{1,1} + n_{1,2} + n_{1,3} \quad N = 38$

Critical z Value - Corresponds to a .2 level of significance and provides at least a 80% level of confidence.

$z := 1.834$

<u>**Multiple Comparisons**</u> - "which populations differ from which others"

- If the difference in mean rank (DMR) between 2 samples is greater than the right-hand side (RHS) of the multiple comparison inequality above, there is a statistically significant difference between the 2 samples.
- The direction of the significant difference is determined by noting for each pair which sample has the larger sample mean.

Comparision of Rating 1 and 2

$DMR_{cpi} := \left|R_{cpi_{1,1}} - R_{cpi_{1,2}}\right| \quad DMR_{cpi} = 6.2$

$DMR_{spi} := \left|R_{spi_{1,1}} - R_{spi_{1,2}}\right| \quad DMR_{spi} = 2.9$

$RHS := z\cdot\sqrt{\frac{N\cdot(N+1)}{12}\cdot\left(\frac{1}{n_{1,1}}+\frac{1}{n_{1,2}}\right)} \quad RHS = 7.442$

For CPI $DMR_{cpi} - RHS = -1.242$ — no significant difference <==

For SPI $DMR_{spi} - RHS = -4.542$ — no significant difference <==

Comparision of Rating 2 and 3

$DMR_{cpi} := \left|R_{cpi_{1,2}} - R_{cpi_{1,3}}\right| \quad DMR_{cpi} = 7.4$

$DMR_{spi} := \left|R_{spi_{1,2}} - R_{spi_{1,3}}\right| \quad DMR_{spi} = 12.2$

$RHS := z\cdot\sqrt{\frac{N\cdot(N+1)}{12}\cdot\left(\frac{1}{n_{1,2}}+\frac{1}{n_{1,3}}\right)} \quad RHS = 8.923$

For CPI $DMR_{cpi} - RHS = -1.523$ — no significant difference <==

For SPI $DMR_{spi} - RHS = 3.277$ — SPI at Rating 3 > SPI at Rating 2 <==

Comparision of Rating 1 and 3

$DMR_{cpi} := \left|R_{cpi_{1,1}} - R_{cpi_{1,3}}\right| \quad DMR_{cpi} = 13.6$

$DMR_{spi} := \left|R_{spi_{1,1}} - R_{spi_{1,3}}\right| \quad DMR_{spi} = 9.3$

$RHS := z\cdot\sqrt{\frac{N\cdot(N+1)}{12}\cdot\left(\frac{1}{n_{1,1}}+\frac{1}{n_{1,3}}\right)} \quad RHS = 8.923$

For CPI $DMR_{cpi} - RHS = 4.677$ — CPI at Rating 3 > CPI at Rating 1 <==

For SPI $DMR_{spi} - RHS = 0.377$ — SPI at Rating 3 > SPI at Rating 1 <==

Figure F-9 Calculations for Multiple Comparison Test for Less than 15% change in Baseline

6. *Multiple Comparison Calculations for Greater Than 15% Change in Baseline*

The Kruskal-Wallis H Test Statistic

$$H := \frac{12}{N\cdot(N+1)}\cdot\sum_{j=1}^{k}\frac{\left[R_j - \frac{n_j\cdot(N+1)}{2}\right]^2}{n_j}$$

Multiple Comparison Inequality

$$\left|\overline{R_i} - \overline{R_j}\right| \le z\cdot\sqrt{\frac{N\cdot(N+1)}{12}\cdot\left(\frac{1}{n_i}+\frac{1}{n_j}\right)}$$

where
k = number of samples
R_i, R_j = sum of the ranks in the ith and jth sample, respectively
n_i, n_j = number of observations in the ith and jth sample, respectively
N = total number of observations

Mean Rank of Ratings 1, 2, & 3 for CPI
$R_{cpi} := (5.5\ \ 10.3\ \ 7.0)$

Mean Rank of Ratings 1, 2, & 3 for SPI
$R_{spi} := (5.0\ \ 6.0\ \ 8.6)$

Number of Observations in Ratings 1, 2, & 3
$n := (2\ \ 3\ \ 9)$

Calculation of N $\quad N := n_{1,1} + n_{1,2} + n_{1,3} \quad N = 14$

Critical z Value - Corresponds to a .2 level of significance and provides at least a 80% level of confidence.
$z := 1.834$

Multiple Comparisons - "which populations differ from which others"
- If the difference in mean rank (DMR) between 2 samples is greater than the right-hand side (RHS) of the multiple comparison inequality above, there is a statistically significant difference between the 2 samples.
- The direction of the significant difference is determined by noting for each pair which sample has the larger sample mean.

Comparision of Rating 1 and 2

$DMR_{cpi} := \left|R_{cpi_{1,1}} - R_{cpi_{1,2}}\right| \quad DMR_{cpi} = 4.8$

$DMR_{spi} := \left|R_{spi_{1,1}} - R_{spi_{1,2}}\right| \quad DMR_{spi} = 1$

$RHS := z\cdot\sqrt{\frac{N\cdot(N+1)}{12}\cdot\left(\frac{1}{n_{1,1}}+\frac{1}{n_{1,2}}\right)} \quad RHS = 7.004$

For CPI $DMR_{cpi} - RHS = -2.204$ no significant difference <==
For SPI $DMR_{spi} - RHS = -6.004$ no significant difference <==

Comparision of Rating 2 and 3

$DMR_{cpi} := \left|R_{cpi_{1,2}} - R_{cpi_{1,3}}\right| \quad DMR_{cpi} = 3.3$

$DMR_{spi} := \left|R_{spi_{1,2}} - R_{spi_{1,3}}\right| \quad DMR_{spi} = 2.6$

$RHS := z\cdot\sqrt{\frac{N\cdot(N+1)}{12}\cdot\left(\frac{1}{n_{1,2}}+\frac{1}{n_{1,3}}\right)} \quad RHS = 5.115$

For CPI $DMR_{cpi} - RHS = -1.815$ no significant difference <==
For SPI $DMR_{spi} - RHS = -2.515$ no significant difference <==

Comparision of Rating 1 and 3

$DMR_{cpi} := \left|R_{cpi_{1,1}} - R_{cpi_{1,3}}\right| \quad DMR_{cpi} = 1.5$

$DMR_{spi} := \left|R_{spi_{1,1}} - R_{spi_{1,3}}\right| \quad DMR_{spi} = 3.6$

$RHS := z\cdot\sqrt{\frac{N\cdot(N+1)}{12}\cdot\left(\frac{1}{n_{1,1}}+\frac{1}{n_{1,3}}\right)} \quad RHS = 5.998$

For CPI $DMR_{cpi} - RHS = -4.498$ no significant difference <==
For SPI $DMR_{spi} - RHS = -2.398$ no significant difference <==

Figure F-10 Calculations for Multiple Comparison Test for Greater than 15% change in Baseline

Appendix G: Data Supporting the Analysis of Moderator "Contract Type"

This appendix contains the complete set of plots, tables, and calculations supporting section 5.2.2. The plots and the Kruskal-Wallis tables for nonparametric analysis of variance are from the Statistix 4.0 computer program. The multiple comparison calculations were performed using Mathcad 4.0. Note that we did not abridge the data as we transcribed it from these computer programs into the report, and thus the number of digits reported in each calculation are not necessarily significant.

The type of contract used to procure systems fundamentally influences the relationship between the Government and the contractor. For example, a fixed-price contract tends to place the monetary risk on the contractor, while a cost-type contract shifts most of the monetary risk to the Government (Nicholas 1990:497). The apportionment of risk between the parties affects how the task is proposed, costed, structured, performed, and tracked. Such a profound environmental moderator may have an effect on the correlation between performance and rating.

1. Scatter Plots of CPI and SPI

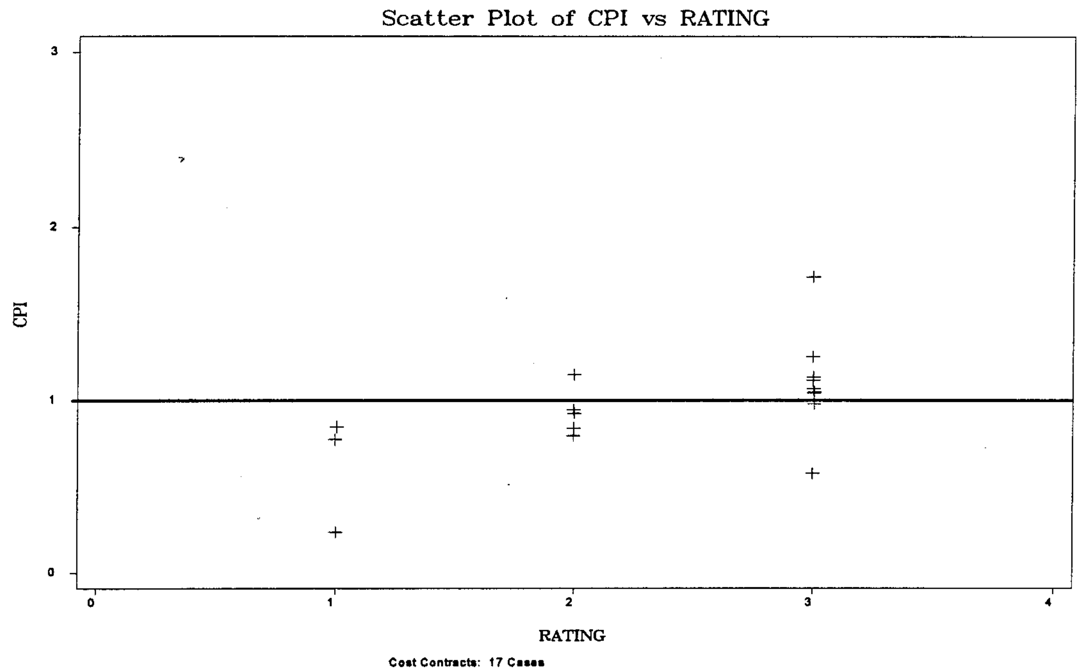

Figure G-1 Scatter Plot of CPI versus Rating for Cost Contracts

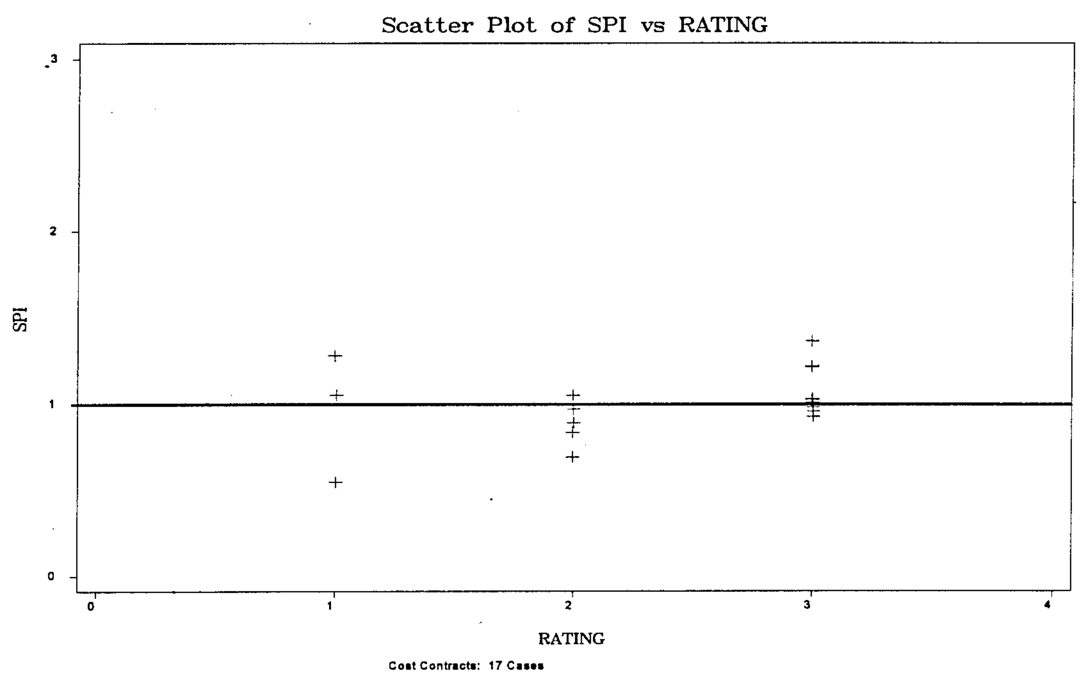

Figure G-2 Scatter Plot of SPI versus Rating for Cost Contracts

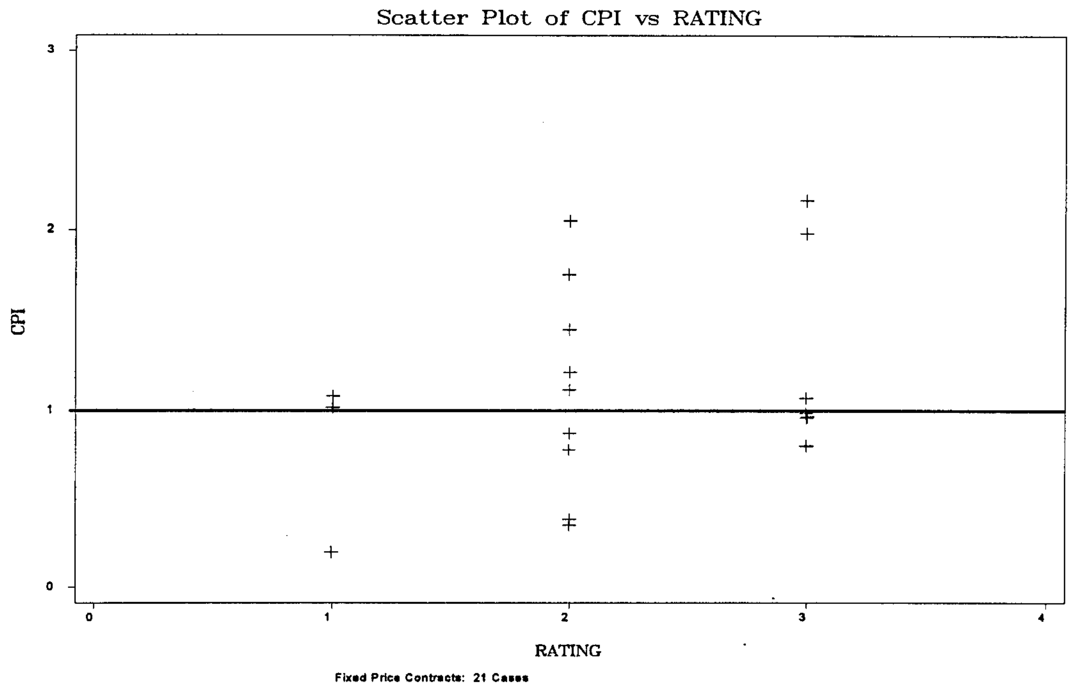

Figure G-3 Scatter Plot of CPI versus Rating for Fixed Price Contracts

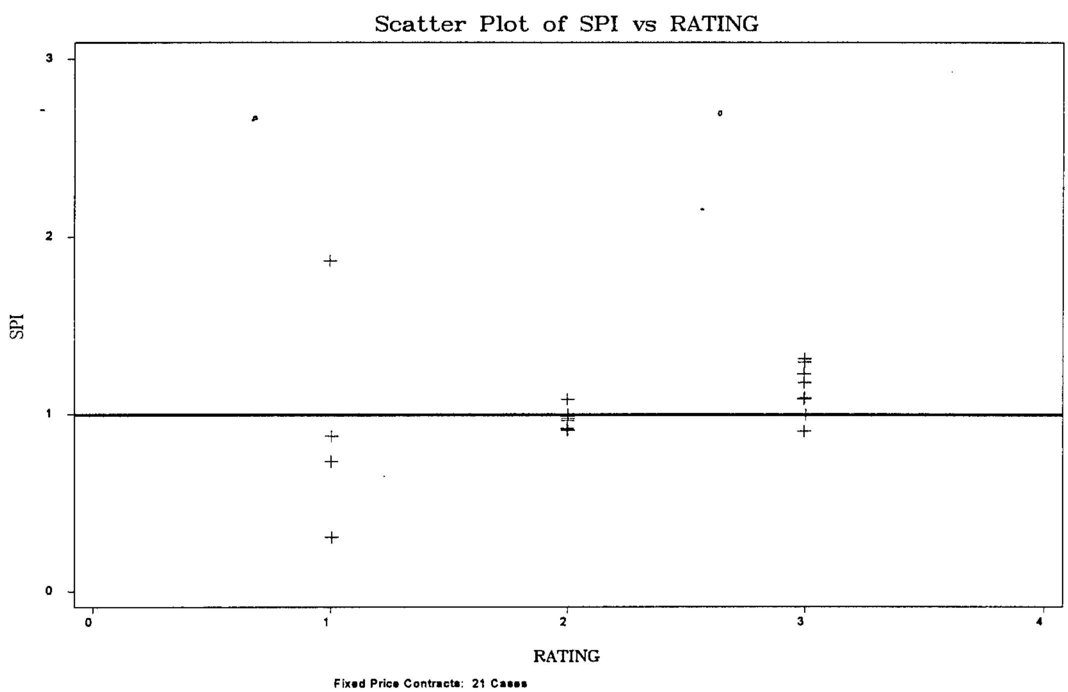

Figure G-4 Scatter Plot of SPI versus Rating for Fixed Price Contracts

2. Box and Whisker Plots of CPI and SPI

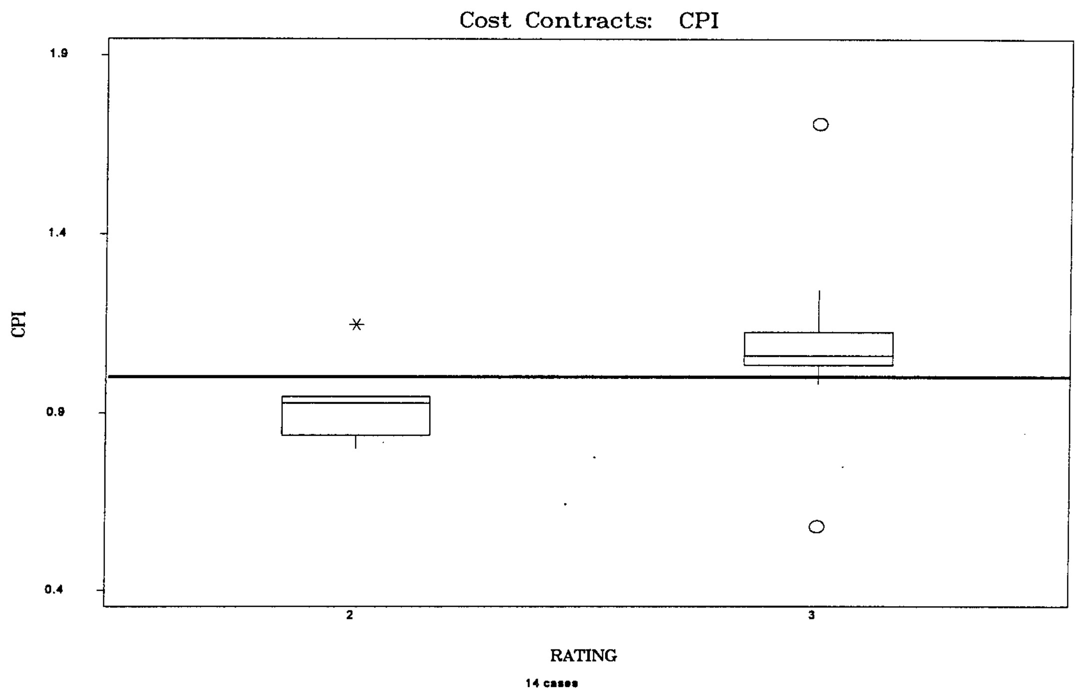

Figure G-5 Box and Whiskers Plot for CPI for Cost Contracts

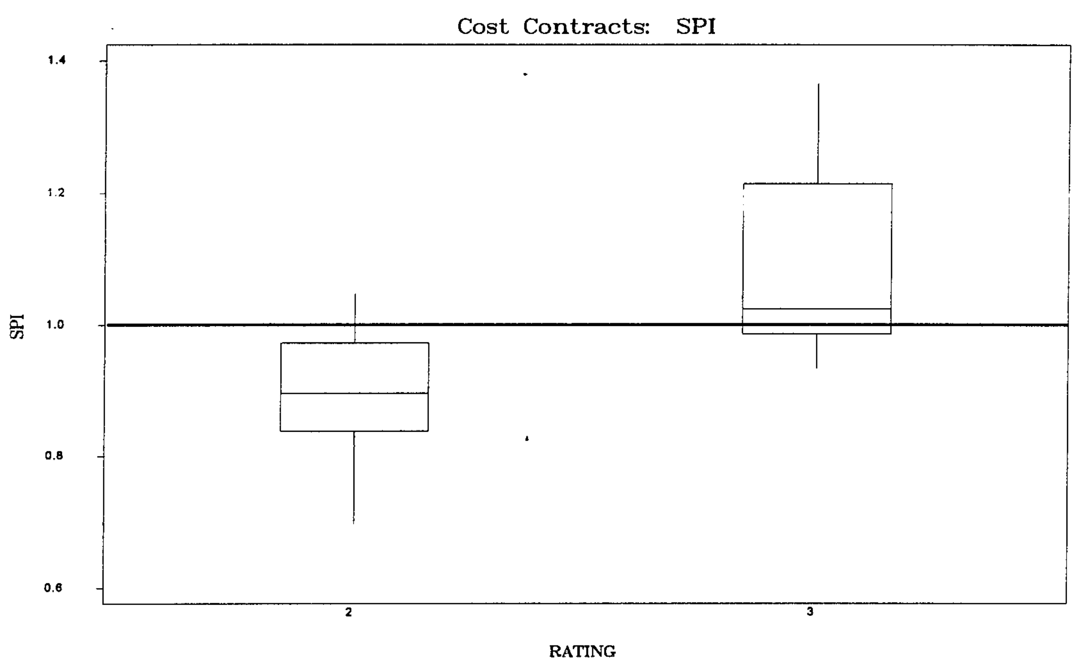

Figure G-6 Box and Whiskers Plot for SPI for Cost Contracts

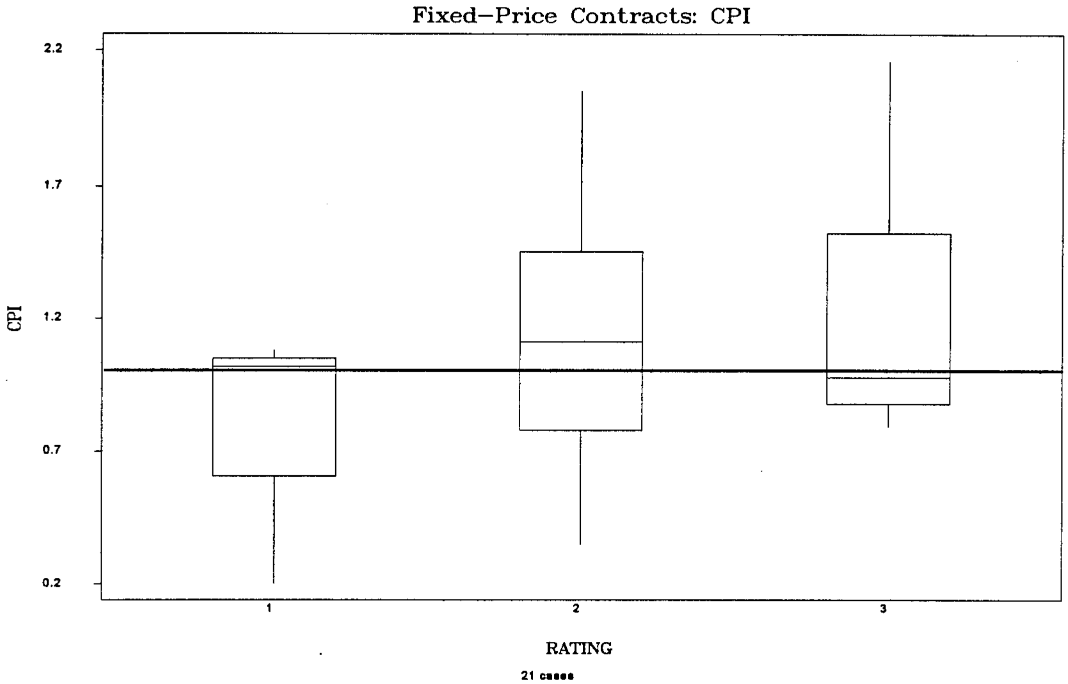

Figure G-7 Box and Whiskers Plot for CPI for Fixed Price Contracts

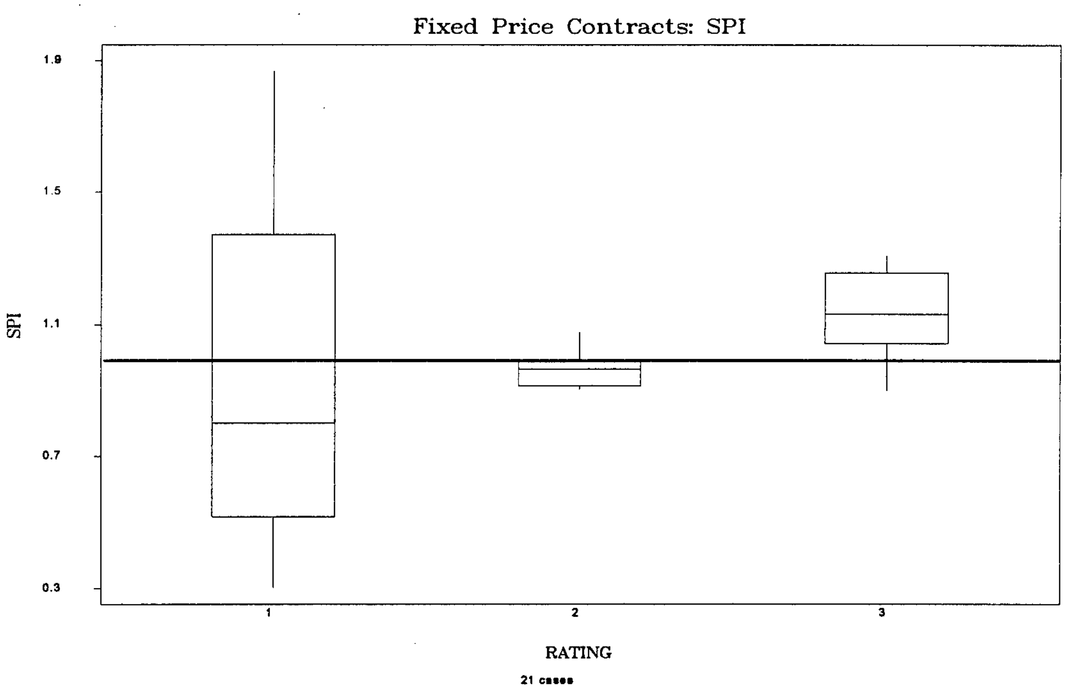

Figure G-8 Box and Whiskers Plot for SPI for Fixed Price Contracts

3. Kruskal-Wallis and Multiple Comparison Tests

Table G-1
Kruskal-Wallis for CPI for Cost Contracts

```
KRUSKAL-WALLIS ONE-WAY NONPARAMETRIC AOV FOR CPI BY RATING

          MEAN   SAMPLE
 RATING   RANK   SIZE
      1    3.3      3
      2    7.8      5
      3   11.6      9
TOTAL      9.0     17

KRUSKAL-WALLIS STATISTIC               6.3651
P-VALUE, USING CHI-SQUARED APPROXIMATION    0.0415
```

Table G-2
Multiple Comparison Matrix for CPI for Cost Contracts

			Rating		
Rating	n	Mean Rank	1	2	3
1	3	3.3		--	--
2	5	7.8	-2.263		--
3	9	11.6	2.126	-1.366	

K-W Statistic of 6.3651, P=0.0415

Table G-3
Kruskal-Wallis for SPI for Cost Contracts

```
KRUSKAL-WALLIS ONE-WAY NONPARAMETRIC AOV FOR SPI BY RATING

          MEAN   SAMPLE
 RATING   RANK   SIZE
      1   10.0      3
      2    5.6      5
      3   10.6      9
TOTAL      9.0     17

KRUSKAL-WALLIS STATISTIC               3.2383
P-VALUE, USING CHI-SQUARED APPROXIMATION    0.1981
```

Table G-4
Multiple Comparison Matrix for SPI for Cost Contracts

			Rating		
Rating	n	Mean Rank	1	2	3
1	3	10.0		--	--
2	5	5.6	-2.363		--
3	9	10.6	-5.574	-0.166	

K-W Statistic of 3.2383, P=0.1981

Table G-5
Kruskal-Wallis for CPI for Fixed Price Contracts

```
KRUSKAL-WALLIS ONE-WAY NONPARAMETRIC AOV FOR CPI BY RATING

          MEAN    SAMPLE
 RATING   RANK    SIZE
      1    9.5       4
      2   11.3       9
      3   11.4       8
TOTAL     11.0      21

KRUSKAL-WALLIS STATISTIC                0.2890
P-VALUE, USING CHI-SQUARED APPROXIMATION      0.8655
```

Table G-6
Multiple Comparison Matrix for CPI for Fixed Price Contracts

			Rating		
Rating	n	Mean Rank	1	2	3
1	4	9.5		--	--
2	9	11.3	-5.038		--
3	8	11.4	-5.069	-5.43	

K-W Statistic of .2890, P=0.8655

Table G-7
Kruskal-Wallis for SPI for Fixed Price Contracts

```
KRUSKAL-WALLIS ONE-WAY NONPARAMETRIC AOV FOR SPI BY RATING

          MEAN    SAMPLE
 RATING   RANK    SIZE
      1    6.8       4
      2    9.2       9
      3   15.2       8
TOTAL     11.0      21

KRUSKAL-WALLIS STATISTIC                6.3101
P-VALUE, USING CHI-SQUARED APPROXIMATION      0.0426
```

Table G-8
Multiple Comparison Matrix for SPI for Fixed Price Contracts

			Rating		
Rating	n	Mean Rank	1	2	3
1	4	6.8		--	--
2	9	9.2	-4.438		--
3	8	15.2	1.431	0.47	

K-W Statistic of 6.3101, P=0.0426

4. Descriptive Statistics

Table G-9
Descriptive Statistics for Cost Contracts

	Rating=1 CPI	Rating=1 SPI	Rating=2 CPI	Rating=2 SPI	Rating=3 CPI	Rating=3 SPI
N	3	3	5	5	9	9
Mean	0.6191	0.9608	0.9316	0.8905	1.1001	1.0811
Std Dev	0.3338	0.3718	0.1354	0.1335	0.2927	0.1483
Min	0.2363	0.5507	0.8000	0.6978	0.5808	0.9335
Median	0.7711	1.0560	0.9273	0.8963	1.0611	1.0242
Max	0.8498	1.2757	1.1479	1.0474	1.7097	1.3652
MAD	0.0787	0.2197	0.0905	0.0766	0.0661	0.0705
Skew	-0.6632	-0.4396	0.7890	-0.3387	0.4619	0.8551

Table G-10
Descriptive Statistics for Fixed Price Contracts

	Rating=1 CPI	Rating=1 SPI	Rating=2 CPI	Rating=2 SPI	Rating=3 CPI	Rating=3 SPI
N	4	4	9	9	8	8
Mean	0.8284	0.9433	1.1081	0.9611	1.2139	1.1338
Std Dev	0.4187	0.6621	0.5812	0.0569	0.5393	0.1419
Min	0.2019	0.3028	0.3496	0.9043	0.7924	0.8998
Median	1.0165	0.8015	1.1114	0.9660	0.9765	1.1322
Max	1.0788	1.8676	2.0506	1.0774	2.1602	1.3073
MAD	0.0322	0.2849	0.3432	0.0474	0.1355	0.1104
Skew	-1.1374	0.6854	0.1955	0.8254	1.0770	-0.2942

5. Multiple Comparison Calculations for Cost Contracts

The Kruskal-Wallis H Test Statistic

$$H := \frac{12}{N \cdot (N+1)} \cdot \sum_{j=1}^{k} \frac{\left[R_j - \frac{n_j \cdot (N+1)}{2}\right]^2}{n_j}$$

Multiple Comparison Inequality

$$\left|\overline{R_i} - \overline{R_j}\right| \le z \cdot \sqrt{\frac{N \cdot (N+1)}{12} \cdot \left(\frac{1}{n_i} + \frac{1}{n_j}\right)}$$

where
k = number of samples
R_i, R_j = sum of the ranks in the ith and jth sample, respectively
n_i, n_j = number of observations in the ith and jth sample, respectively
N = total number of observations

Mean Rank of Ratings 1, 2, & 3 for CPI
$R_{cpi} := (3.3\ \ 7.8\ \ 11.6)$

Mean Rank of Ratings 1, 2, & 3 for SPI
$R_{spi} := (10.0\ \ 5.6\ \ 10.6)$

Number of Observations in Ratings 1, 2, & 3
$n := (3\ \ 5\ \ 9)$

Calculation of N $\quad N := n_{1,1} + n_{1,2} + n_{1,3} \quad N = 17$

Critical z Value - Corresponds to a .2 level of significance and provides at least a 80% level of confidence.
$z := 1.834$

<u>Multiple Comparisons</u> - "which populations differ from which others"
- If the difference in mean rank (DMR) between 2 samples is greater than the right-hand side (RHS) of the multiple comparison inequality above, there is a statistically significant difference between the 2 samples.
- The direction of the significant difference is determined by noting for each pair which sample has the larger sample mean.

Comparision of Rating 1 and 2

$DMR_{cpi} := \left|R_{cpi_{1,1}} - R_{cpi_{1,2}}\right| \quad DMR_{cpi} = 4.5$

$DMR_{spi} := \left|R_{spi_{1,1}} - R_{spi_{1,2}}\right| \quad DMR_{spi} = 4.4$

$RHS := z \cdot \sqrt{\frac{N \cdot (N+1)}{12} \cdot \left(\frac{1}{n_{1,1}} + \frac{1}{n_{1,2}}\right)} \quad RHS = 6.763$

For CPI $DMR_{cpi} - RHS = -2.263$ — no significant difference <===
For SPI $DMR_{spi} - RHS = -2.363$ — no significant difference <===

Comparision of Rating 2 and 3

$DMR_{cpi} := \left|R_{cpi_{1,2}} - R_{cpi_{1,3}}\right| \quad DMR_{cpi} = 3.8$

$DMR_{spi} := \left|R_{spi_{1,2}} - R_{spi_{1,3}}\right| \quad DMR_{spi} = 5$

$RHS := z \cdot \sqrt{\frac{N \cdot (N+1)}{12} \cdot \left(\frac{1}{n_{1,2}} + \frac{1}{n_{1,3}}\right)} \quad RHS = 5.166$

For CPI $DMR_{cpi} - RHS = -1.366$ — no significant difference <===
For SPI $DMR_{spi} - RHS = -0.166$ — no significant difference <===

Comparision of Rating 1 and 3

$DMR_{cpi} := \left|R_{cpi_{1,1}} - R_{cpi_{1,3}}\right| \quad DMR_{cpi} = 8.3$

$DMR_{spi} := \left|R_{spi_{1,1}} - R_{spi_{1,3}}\right| \quad DMR_{spi} = 0.6$

$RHS := z \cdot \sqrt{\frac{N \cdot (N+1)}{12} \cdot \left(\frac{1}{n_{1,1}} + \frac{1}{n_{1,3}}\right)} \quad RHS = 6.174$

For CPI $DMR_{cpi} - RHS = 2.126$ — CPI at Rating 3 > CPI at Rating 1 <===
For SPI $DMR_{spi} - RHS = -5.574$ — no significant difference <===

Figure G-9 Calculations for Multiple Comparison Test for Cost Contracts

6. Multiple Comparison Calculations for Fixed Price Contracts

The Kruskal-Wallis H Test Statistic

$$H := \frac{12}{N \cdot (N+1)} \cdot \sum_{j=1}^{k} \frac{\left[R_j - \frac{n_j \cdot (N+1)}{2}\right]^2}{n_j}$$

Multiple Comparison Inequality

$$\left|\overline{R_i} - \overline{R_j}\right| \le z \cdot \sqrt{\frac{N \cdot (N+1)}{12} \cdot \left(\frac{1}{n_i} + \frac{1}{n_j}\right)}$$

where
k = number of samples
R_i, R_j = sum of the ranks in the ith and jth sample, respectively
n_i, n_j = number of observations in the ith and jth sample, respectively
N = total number of observations

Mean Rank of Ratings 1, 2, & 3 for CPI
$R_{cpi} := (9.5\ \ 11.3\ \ 11.4)$

Mean Rank of Ratings 1, 2, & 3 for SPI
$R_{spi} := (6.8\ \ 9.2\ \ 15.2)$

Number of Observations in Ratings 1, 2, & 3
$n := (4\ \ 9\ \ 8)$

Calculation of N $\quad N := n_{1,1} + n_{1,2} + n_{1,3} \qquad N = 21$

Critical z Value - Corresponds to a .2 level of significance and provides at least a 80% level of confidence.
$z := 1.834$

<u>**Multiple Comparisons**</u> - "which populations differ from which others"
- If the difference in mean rank (DMR) between 2 samples is greater than the right-hand side (RHS) of the multiple comparison inequality above, there is a statistically significant difference between the 2 samples.
- The direction of the significant difference is determined by noting for each pair which sample has the larger sample mean.

Comparision of Rating 1 and 2

$DMR_{cpi} := \left|R_{cpi_{1,1}} - R_{cpi_{1,2}}\right| \qquad DMR_{cpi} = 1.8$

$DMR_{spi} := \left|R_{spi_{1,1}} - R_{spi_{1,2}}\right| \qquad DMR_{spi} = 2.4$

$RHS := z \cdot \sqrt{\frac{N \cdot (N+1)}{12} \cdot \left(\frac{1}{n_{1,1}} + \frac{1}{n_{1,2}}\right)} \qquad RHS = 6.838$

For CPI $DMR_{cpi} - RHS = -5.038$ — no significant difference <==

For SPI $DMR_{spi} - RHS = -4.438$ — no significant difference <==

Comparision of Rating 2 and 3

$DMR_{cpi} := \left|R_{cpi_{1,2}} - R_{cpi_{1,3}}\right| \qquad DMR_{cpi} = 0.1$

$DMR_{spi} := \left|R_{spi_{1,2}} - R_{spi_{1,3}}\right| \qquad DMR_{spi} = 6$

$RHS := z \cdot \sqrt{\frac{N \cdot (N+1)}{12} \cdot \left(\frac{1}{n_{1,2}} + \frac{1}{n_{1,3}}\right)} \qquad RHS = 5.53$

For CPI $DMR_{cpi} - RHS = -5.43$ — no significant difference <==

For SPI $DMR_{spi} - RHS = 0.47$ — SPI at Rating 3 > SPI at Rating 2 <==

Comparision of Rating 1 and 3

$DMR_{cpi} := \left|R_{cpi_{1,1}} - R_{cpi_{1,3}}\right| \qquad DMR_{cpi} = 1.9$

$DMR_{spi} := \left|R_{spi_{1,1}} - R_{spi_{1,3}}\right| \qquad DMR_{spi} = 8.4$

$RHS := z \cdot \sqrt{\frac{N \cdot (N+1)}{12} \cdot \left(\frac{1}{n_{1,1}} + \frac{1}{n_{1,3}}\right)} \qquad RHS = 6.969$

For CPI $DMR_{cpi} - RHS = -5.069$ — no significant difference <==

For SPI $DMR_{spi} - RHS = 1.431$ — SPI at Rating 3 > SPI at Rating 1 <==

Figure G-10 Calculations for Multiple Comparison Test for Fixed Price Contracts

Appendix H: Data Supporting the Analysis of Moderator "Percent Complete"

This appendix contains the complete set of plots, tables, and calculations supporting section 5.2.2. The plots and the Kruskal-Wallis tables for nonparametric analysis of variance are from the Statistix 4.0 computer program. The multiple comparison calculations were performed using Mathcad 4.0. Note that we did not abridge the data as we transcribed it from these computer programs into the report, and thus the number of digits reported in each calculation are not necessarily significant.

In our review of the literature, we found that proximity to completion has a significant effect on the dynamics of the cumulative performance indices. For example, cumulative SPI, by definition, is driven to 1.00 at program completion while cumulative CPI has been shown to be stable from the 20% completion point, where "stability" is defined as CPI range being less than 0.2. The dynamics of the cumulative performance indices have been well noted in the literature, and are a fundamental element in the art of estimating at-complete costs (Christensen and Heise 1993:7-15) In our research, however, we are taking a 12-month slice of these performance indices. We acknowledge these "snapshot" indices will not be as stable as the cumulative indices. Nevertheless, it was important that our research capture the degree to which the dynamics of the cumulative indices affected our non-cumulative indices.

For our dataset, we chose 80 percent complete as the point about which we stratified the sample. This was done to distinguish between the performance over the bulk of the contract and the performance near program completion.

1. Scatter Plots of CPI and SPI

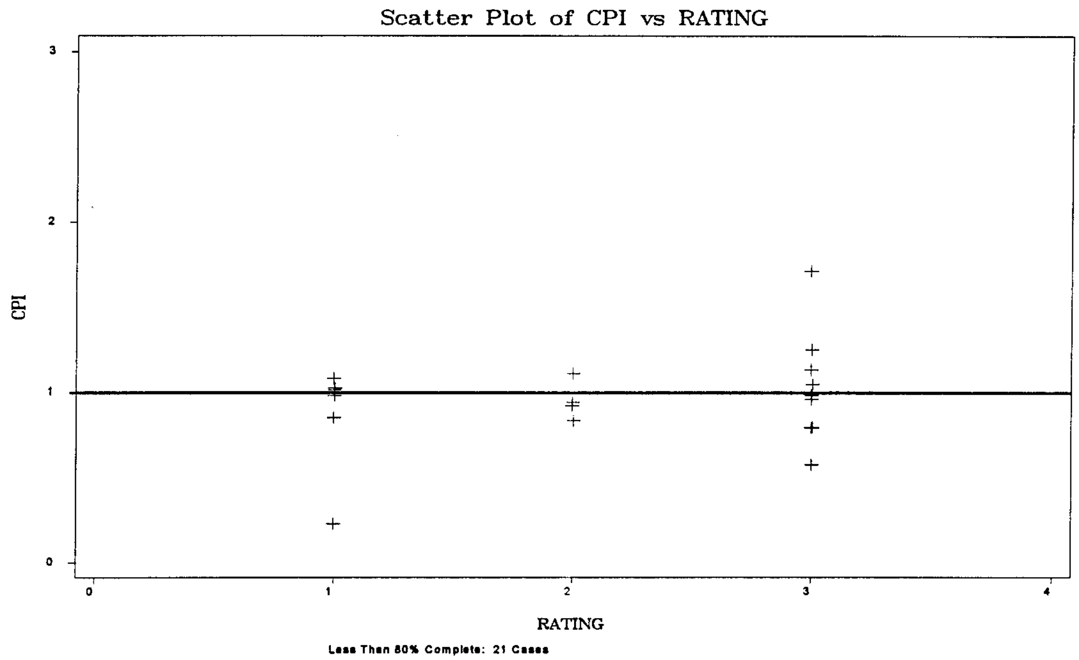

Figure H-1 Scatter Plot of CPI versus Rating for Less than 80% Complete

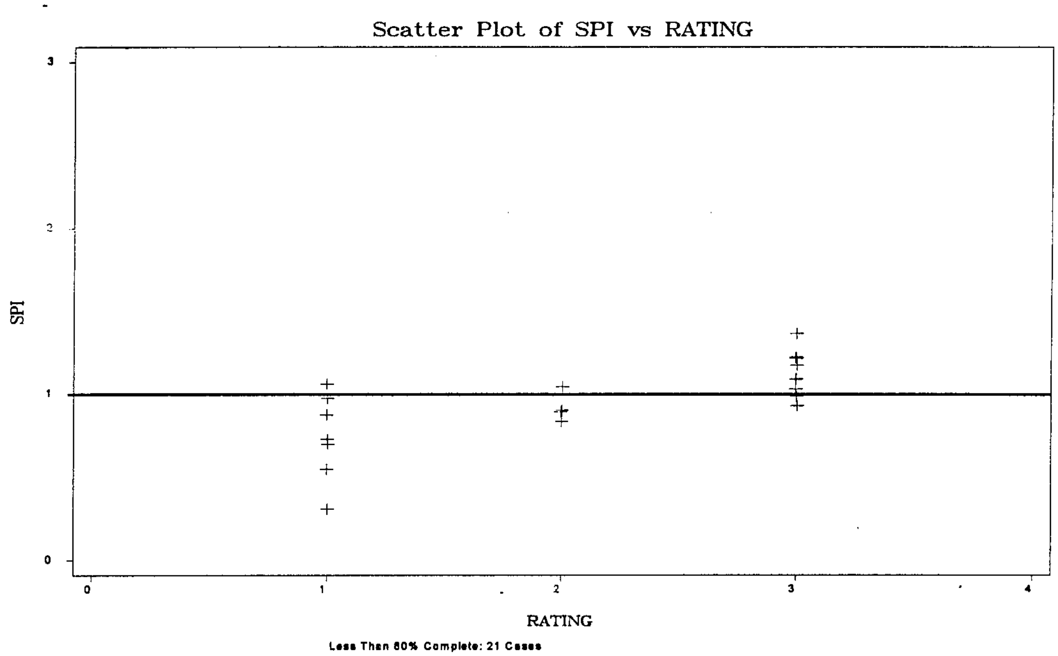

Figure H-2 Scatter Plot of SPI versus Rating for Less than 80% Complete

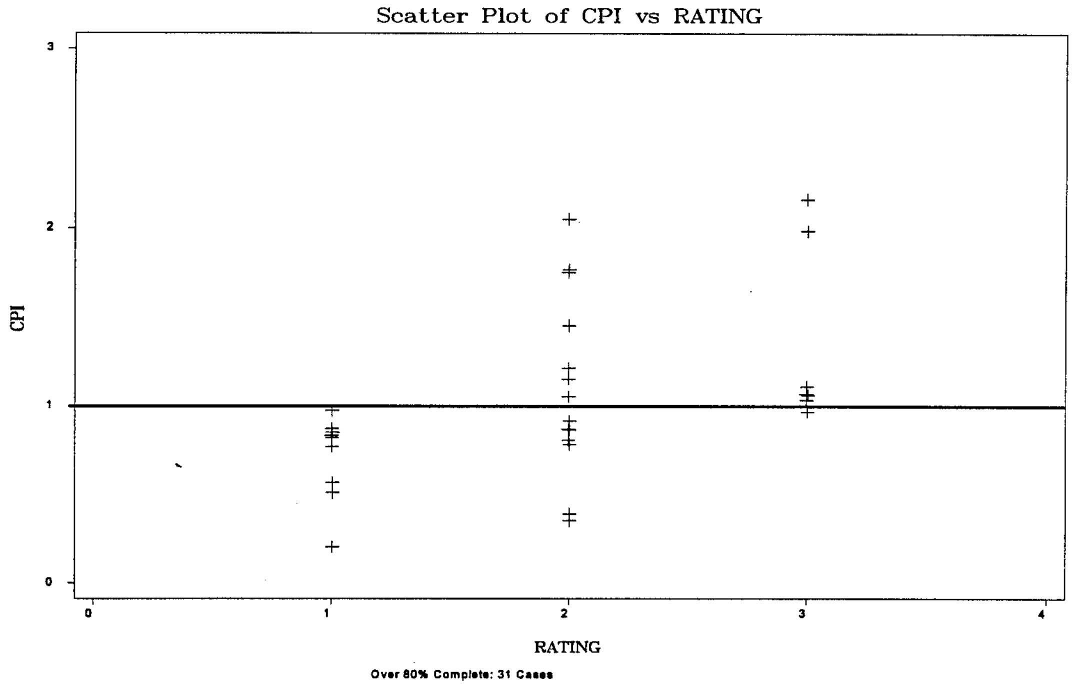

Figure H-3 Scatter Plot of CPI versus Rating for Greater than 80% Complete

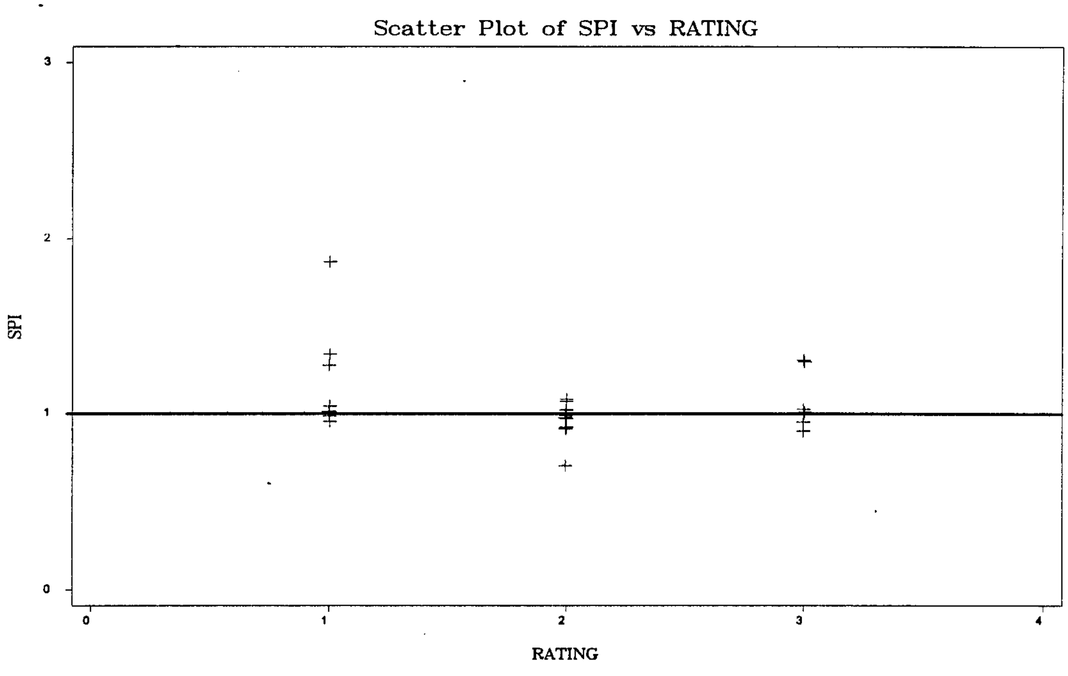

Figure H-4 Scatter Plot of SPI versus Rating for Greater than 80% Complete

2. *Box and Whisker Plots of CPI and SPI*

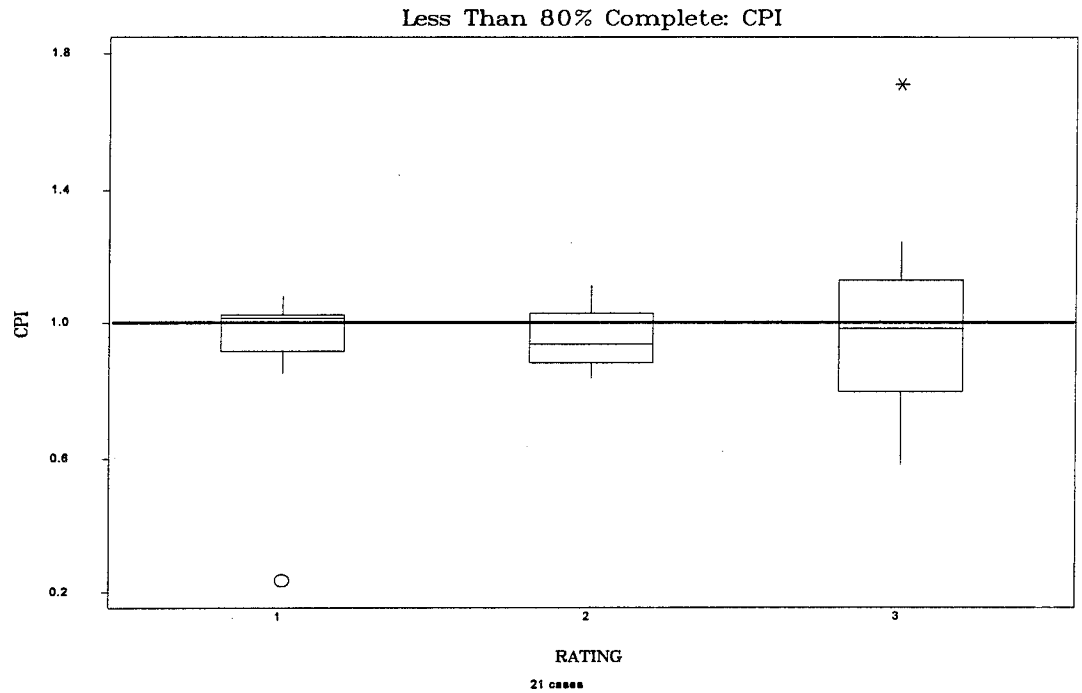

Figure H-5 Box and Whiskers Plot of CPI for Less than 80% Complete

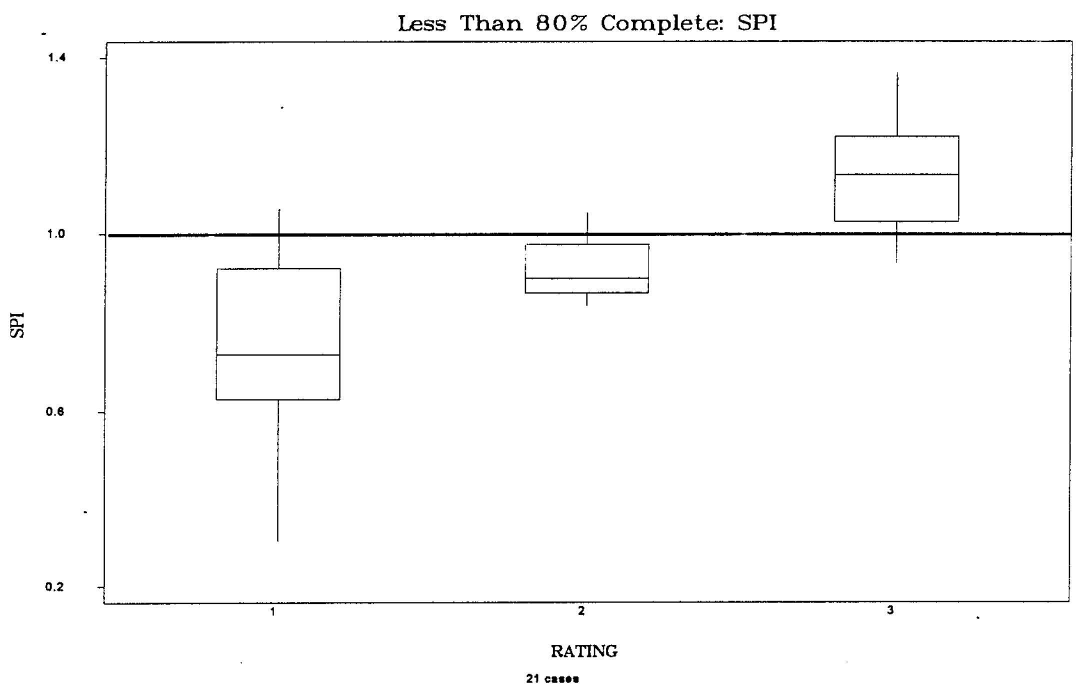

Figure H-6 Box and Whiskers Plot of SPI for Less than 80% Complete

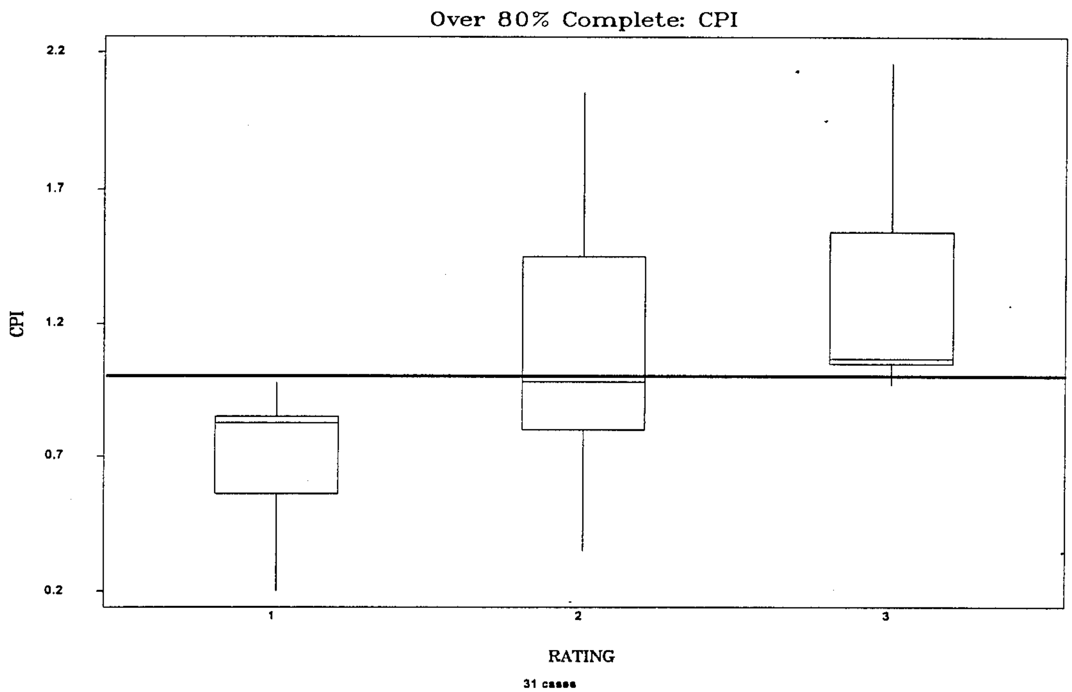

Figure H-7 Box and Whiskers Plot of CPI for Greater than 80% Complete

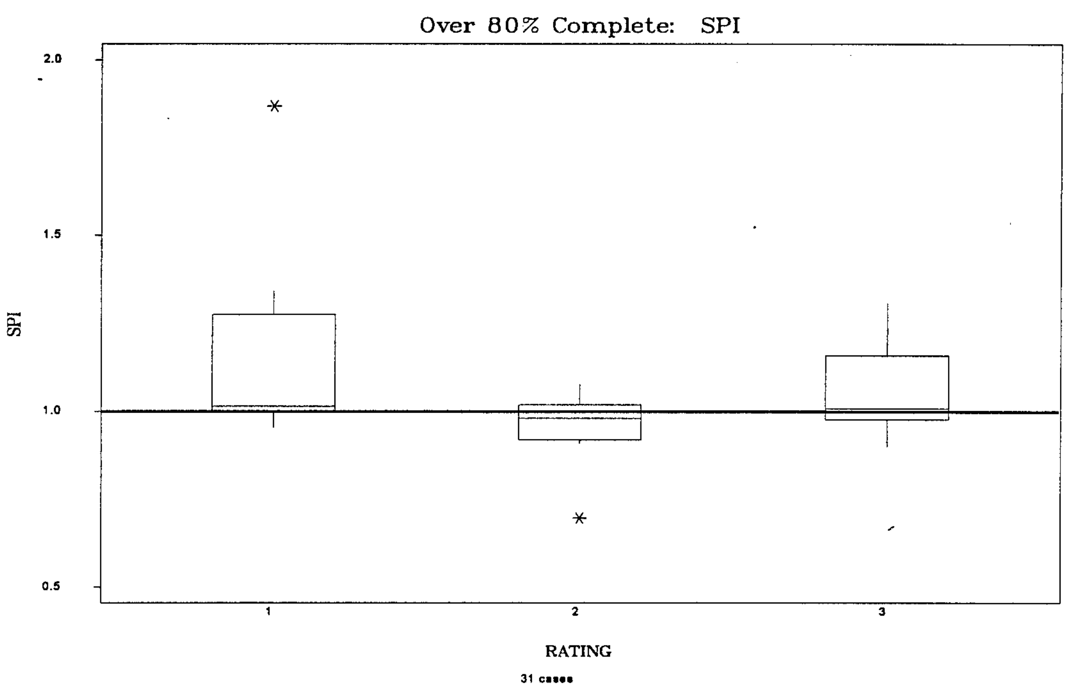

Figure H-8 Box and Whiskers Plot of SPI for Greater than 80% Complete

3. Kruskal-Wallis and Multiple Comparison Tests

Table H-1
Kruskal-Wallis for CPI for Less than 80% Complete

```
KRUSKAL-WALLIS ONE-WAY NONPARAMETRIC AOV FOR CPI BY RATING

         MEAN   SAMPLE
RATING   RANK   SIZE
     1   10.9      7
     2    9.5      4
     3   11.7     10
TOTAL    11.0     21

KRUSKAL-WALLIS STATISTIC                 0.3647
P-VALUE, USING CHI-SQUARED APPROXIMATION     0.8333
```

Table H-2
Multiple Comparison Matrix for CPI for Less than 80% Complete

			Rating		
Rating	n	Mean Rank	1	2	3
1	7	10.9		--	--
2	4	9.5	-5.733		--
3	10	11.7	-4.808	-4.532	

K-W Statistic of .3647, P=0.8333

Table H-3
Kruskal-Wallis for SPI for Less than 80% Complete

```
KRUSKAL-WALLIS ONE-WAY NONPARAMETRIC AOV FOR SPI BY RATING

         MEAN   SAMPLE
RATING   RANK   SIZE
     1    5.7      7
     2    8.3      4
     3   15.8     10
TOTAL    11.0     21

KRUSKAL-WALLIS STATISTIC                 11.8499
P-VALUE, USING CHI-SQUARED APPROXIMATION     0.0027
```

Table H-4
Multiple Comparison Matrix for SPI for Less than 80% Complete

			Rating		
Rating	n	Mean Rank	1	2	3
1	7	5.7		--	--
2	4	8.3	-4.533		--
3	10	15.8	4.492	0.768	

K-W Statistic of 11.8499, P=0.0027

Table H-5
Kruskal-Wallis for CPI for Greater than 80% Complete

```
KRUSKAL-WALLIS ONE-WAY NONPARAMETRIC AOV FOR CPI BY RATING

          MEAN   SAMPLE
 RATING   RANK    SIZE
      1    9.1      10
      2   17.4      14
      3   23.1       7
TOTAL     16.0      31

KRUSKAL-WALLIS STATISTIC              10.3915
P-VALUE, USING CHI-SQUARED APPROXIMATION    0.0055
```

Table H-6
Multiple Comparison Matrix for CPI for Greater than 80% Complete

			Rating		
Rating	n	Mean Rank	1	2	3
1	10	9.1		--	--
2	14	17.4	1.396		--
3	7	23.1	5.782	-2.019	

K-W Statistic of 10.3915, P=0.0055

Table H-7
Kruskal-Wallis for SPI for Greater than 80% Complete

```
KRUSKAL-WALLIS ONE-WAY NONPARAMETRIC AOV FOR SPI BY RATING

          MEAN   SAMPLE
 RATING   RANK    SIZE
      1   20.1      10
      2   12.5      14
      3   17.2       7
TOTAL     16.0      31

KRUSKAL-WALLIS STATISTIC              4.1921
P-VALUE, USING CHI-SQUARED APPROXIMATION    0.1229
```

Table H-8
Multiple Comparison Matrix for SPI for Greater than 80% Complete

			Rating		
Rating	n	Mean Rank	1	2	3
1	10	20.1		--	--
2	14	12.5	0.696		--
3	7	17.2	-5.318	-3.019	

K-W Statistic of 4.1921, P=0.1229

4. Descriptive Statistics

Table H-9
Descriptive Statistics for Less than 80% Complete

	Rating=1 CPI	Rating=1 SPI	Rating=2 CPI	Rating=2 SPI	Rating=3 CPI	Rating=3 SPI
N	7	7	4	4	10	10
Mean	0.8869	0.7413	0.9553	0.9215	1.0229	1.1316
Std Dev	0.2956	0.2582	0.1144	0.0889	0.3054	0.1306
Min	0.2363	0.3028	0.8368	0.8383	0.5808	0.9335
Median	1.0144	0.7304	0.9365	0.9003	0.9839	1.1322
Max	1.0788	1.0560	1.1114	1.0474	1.7097	1.3652
MAD	0.0325	0.1797	0.0544	0.0330	0.1654	0.0874
Skew	-1.8149	-0.4767	0.5437	0.7634	0.9270	0.1378

Table H-10
Descriptive Statistics for Greater than 80% Complete

	Rating=1 CPI	Rating=1 SPI	Rating=2 CPI	Rating=2 SPI	Rating=3 CPI	Rating=3 SPI
N	10	10	14	14	7	7
Mean	0.7236	1.1498	1.1008	0.9661	1.3405	1.0692
Std Dev	0.2310	0.2841	0.5069	0.0929	0.5035	0.1628
Min	0.2019	0.9535	0.3496	0.6978	0.9674	0.8998
Median	0.8242	1.0143	0.9808	0.9803	1.0675	1.0072
Max	0.9770	1.8676	2.0506	1.0774	2.1602	1.3073
MAD	0.0514	0.0279	0.2194	0.0387	0.0430	0.0535
Skew	-1.2353	1.8100	0.3716	-1.6724	0.9585	0.7304

5. Multiple Comparison Calculations for Less Than 80% Complete

The Kruskal-Wallis H Test Statistic

$$H := \frac{12}{N\cdot(N+1)}\cdot\sum_{j=1}^{k}\frac{\left[R_j - \frac{n_j\cdot(N+1)}{2}\right]^2}{n_j}$$

Multiple Comparison Inequality

$$\left|\overline{R_i} - \overline{R_j}\right| \le z\cdot\sqrt{\frac{N\cdot(N+1)}{12}\cdot\left(\frac{1}{n_i}+\frac{1}{n_j}\right)}$$

where
k = number of samples
R_i, R_j = sum of the ranks in the ith and jth sample, respectively
n_i, n_j = number of observations in the ith and jth sample, respectively
N = total number of observations

Mean Rank of Ratings 1, 2, & 3 for CPI
$R_{cpi} := (10.9\ \ 9.5\ \ 11.7)$

Mean Rank of Ratings 1, 2, & 3 for SPI
$R_{spi} := (5.7\ \ 8.3\ \ 15.8)$

Number of Observations in Ratings 1, 2, & 3
$n := (7\ \ 4\ \ 10)$

Calculation of N $\quad N := n_{1,1} + n_{1,2} + n_{1,3} \quad N = 21$

Critical z Value - Corresponds to a .2 level of significance and provides at least a 80% level of confidence.
$z := 1.834$

<u>**Multiple Comparisons**</u> - "which populations differ from which others"
- If the difference in mean rank (DMR) between 2 samples is greater than the right-hand side (RHS) of the multiple comparison inequality above, there is a statistically significant difference between the 2 samples.
- The direction of the significant difference is determined by noting for each pair which sample has the larger sample mean.

Comparision of Rating 1 and 2

$DMR_{cpi} := \left|R_{cpi_{1,1}} - R_{cpi_{1,2}}\right| \quad DMR_{cpi} = 1.4$

$DMR_{spi} := \left|R_{spi_{1,1}} - R_{spi_{1,2}}\right| \quad DMR_{spi} = 2.6$

$RHS := z\cdot\sqrt{\frac{N\cdot(N+1)}{12}\cdot\left(\frac{1}{n_{1,1}}+\frac{1}{n_{1,2}}\right)} \quad RHS = 7.133$

For CPI $DMR_{cpi} - RHS = -5.733$ — no significant difference <==

For SPI $DMR_{spi} - RHS = -4.533$ — no significant difference <==

Comparision of Rating 2 and 3

$DMR_{cpi} := \left|R_{cpi_{1,2}} - R_{cpi_{1,3}}\right| \quad DMR_{cpi} = 2.2$

$DMR_{spi} := \left|R_{spi_{1,2}} - R_{spi_{1,3}}\right| \quad DMR_{spi} = 7.5$

$RHS := z\cdot\sqrt{\frac{N\cdot(N+1)}{12}\cdot\left(\frac{1}{n_{1,2}}+\frac{1}{n_{1,3}}\right)} \quad RHS = 6.732$

For CPI $DMR_{cpi} - RHS = -4.532$ — no significant difference <==

For SPI $DMR_{spi} - RHS = 0.768$ — SPI at Rating 3 > SPI at Rating 2 <==

Comparision of Rating 1 and 3

$DMR_{cpi} := \left|R_{cpi_{1,1}} - R_{cpi_{1,3}}\right| \quad DMR_{cpi} = 0.8$

$DMR_{spi} := \left|R_{spi_{1,1}} - R_{spi_{1,3}}\right| \quad DMR_{spi} = 10.1$

$RHS := z\cdot\sqrt{\frac{N\cdot(N+1)}{12}\cdot\left(\frac{1}{n_{1,1}}+\frac{1}{n_{1,3}}\right)} \quad RHS = 5.608$

For CPI $DMR_{cpi} - RHS = -4.808$ — no significant difference <==

For SPI $DMR_{spi} - RHS = 4.492$ — SPI at Rating 3 > SPI at Rating 1 <==

Figure H-9 Calculations for Multiple Comparison Test for Less than 80% Complete

6. Multiple Comparison Calculations for Greater Than 80% Complete

The Kruskal-Wallis H Test Statistic

$$H := \frac{12}{N\cdot(N+1)} \cdot \sum_{j=1}^{k} \frac{\left[R_j - \frac{n_j\cdot(N+1)}{2}\right]^2}{n_j}$$

Multiple Comparison Inequality

$$\left|\overline{R_i} - \overline{R_j}\right| \le z\cdot\sqrt{\frac{N\cdot(N+1)}{12}\cdot\left(\frac{1}{n_i}+\frac{1}{n_j}\right)}$$

where
k = number of samples
R_i, R_j = sum of the ranks in the ith and jth sample, respectively
n_i, n_j = number of observations in the ith and jth sample, respectively
N = total number of observations

Mean Rank of Ratings 1, 2, & 3 for CPI
$R_{cpi} := (9.1\ \ 17.4\ \ 23.1)$

Mean Rank of Ratings 1, 2, & 3 for SPI
$R_{spi} := (20.1\ \ 12.5\ \ 17.2)$

Number of Observations in Ratings 1, 2, & 3
$n := (10\ \ 14\ \ 7)$

Calculation of N $\quad N := n_{1,1} + n_{1,2} + n_{1,3} \quad N = 31$

Critical z Value - Corresponds to a .2 level of significance and provides at least a 80% level of confidence.
$z := 1.834$

Multiple Comparisons - "which populations differ from which others"
- If the difference in mean rank (DMR) between 2 samples is greater than the right-hand side (RHS) of the multiple comparison inequality above, there is a statistically significant difference between the 2 samples.
- The direction of the significant difference is determined by noting for each pair which sample has the larger sample mean.

Comparision of Rating 1 and 2

$DMR_{cpi} := \left|R_{cpi_{1,1}} - R_{cpi_{1,2}}\right| \quad DMR_{cpi} = 8.3$

$DMR_{spi} := \left|R_{spi_{1,1}} - R_{spi_{1,2}}\right| \quad DMR_{spi} = 7.6$

$RHS := z\cdot\sqrt{\frac{N\cdot(N+1)}{12}\cdot\left(\frac{1}{n_{1,1}}+\frac{1}{n_{1,2}}\right)} \quad RHS = 6.904$

For CPI $DMR_{cpi} - RHS = 1.396$ — CPI at Rating 2 > CPI at Rating 1 <==
For SPI $DMR_{spi} - RHS = 0.696$ — SPI at Rating 1 > SPI at Rating 2 <==

Comparision of Rating 2 and 3

$DMR_{cpi} := \left|R_{cpi_{1,2}} - R_{cpi_{1,3}}\right| \quad DMR_{cpi} = 5.7$

$DMR_{spi} := \left|R_{spi_{1,2}} - R_{spi_{1,3}}\right| \quad DMR_{spi} = 4.7$

$RHS := z\cdot\sqrt{\frac{N\cdot(N+1)}{12}\cdot\left(\frac{1}{n_{1,2}}+\frac{1}{n_{1,3}}\right)} \quad RHS = 7.719$

For CPI $DMR_{cpi} - RHS = -2.019$ — no significant difference <==
For SPI $DMR_{spi} - RHS = -3.019$ — no significant difference <==

Comparision of Rating 1 and 3

$DMR_{cpi} := \left|R_{cpi_{1,1}} - R_{cpi_{1,3}}\right| \quad DMR_{cpi} = 14$

$DMR_{spi} := \left|R_{spi_{1,1}} - R_{spi_{1,3}}\right| \quad DMR_{spi} = 2.9$

$RHS := z\cdot\sqrt{\frac{N\cdot(N+1)}{12}\cdot\left(\frac{1}{n_{1,1}}+\frac{1}{n_{1,3}}\right)} \quad RHS = 8.218$

For CPI $DMR_{cpi} - RHS = 5.782$ — CPI at Rating 3 > CPI at Rating 1 <==
For SPI $DMR_{spi} - RHS = -5.318$ — no significant difference <==

Figure H-10 Calculations for Multiple Comparison Test for Greater than 80% Complete

Appendix I: Data Supporting the Analysis of Moderator "Application Type"

This appendix contains the complete set of plots, tables, and calculations supporting section 5.2.2. The plots and the Kruskal-Wallis tables for nonparametric analysis of variance are from the Statistix 4.0 computer program. The multiple comparison calculations were performed using Mathcad 4.0. Note that we did not abridge the data as we transcribed it from these computer programs into the report, and thus the number of digits reported in each calculation are not necessarily significant.

Application type is a gross predictor of project complexity. The categories selected, real-time applications versus information systems applications, capture the distinction between the highly complex avionics, flight control, simulation, and command and control applications and the usually less-demanding database and catalog applications.

1. Scatter Plots of CPI and SPI

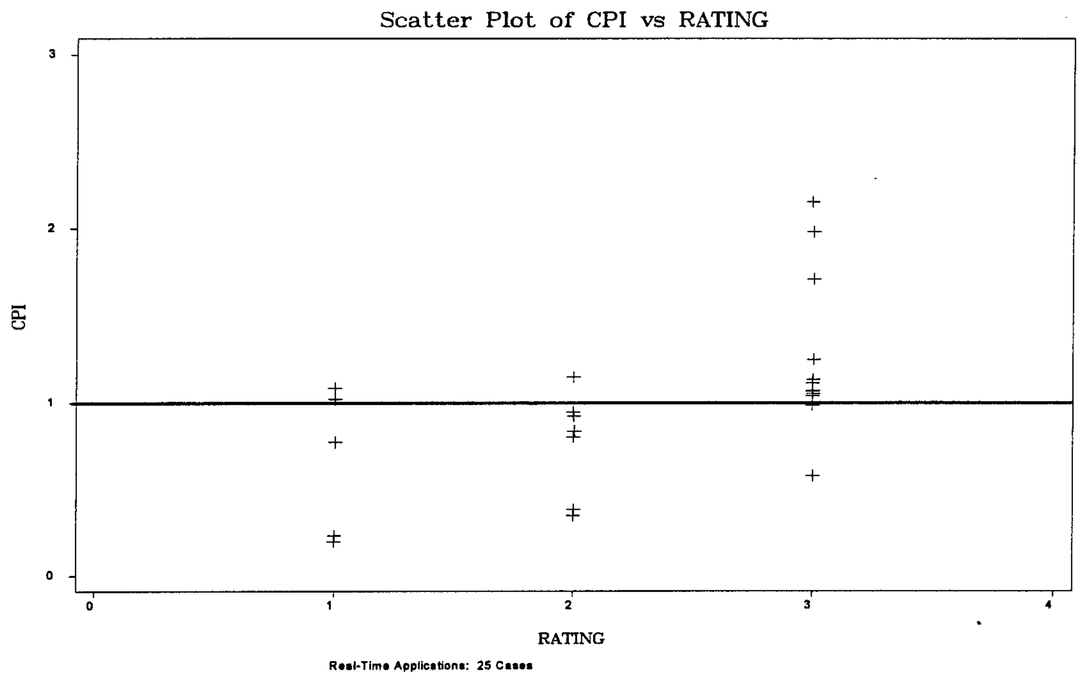

Figure I-1 Scatter Plot of CPI versus Rating for Real-Time Applications

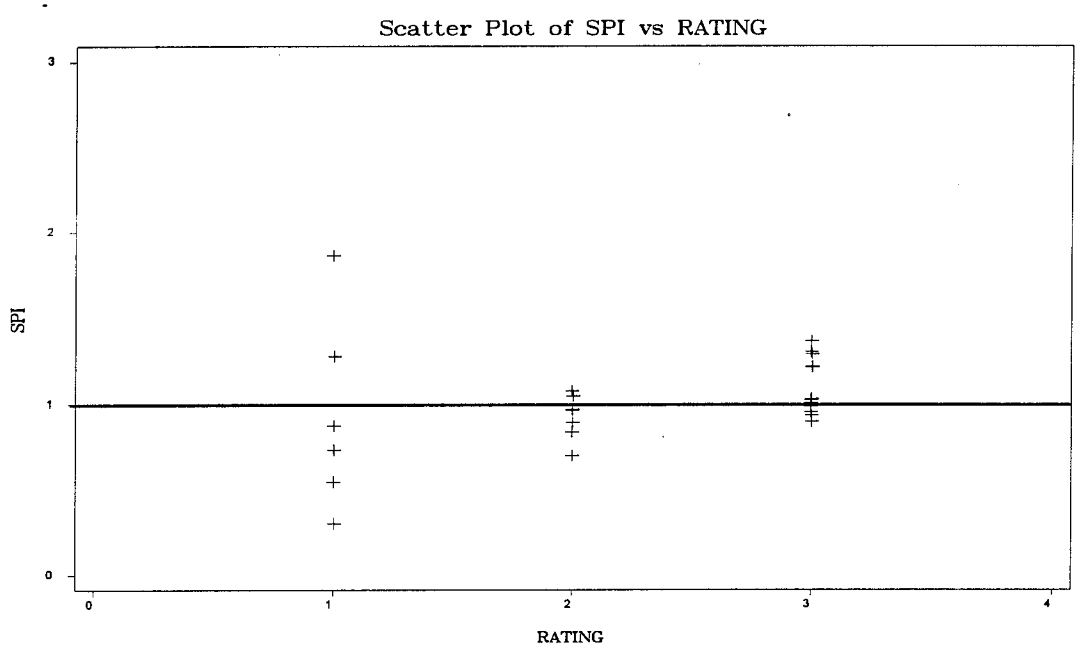

Figure I-2 Scatter Plot of SPI versus Rating for Real-Time Applications

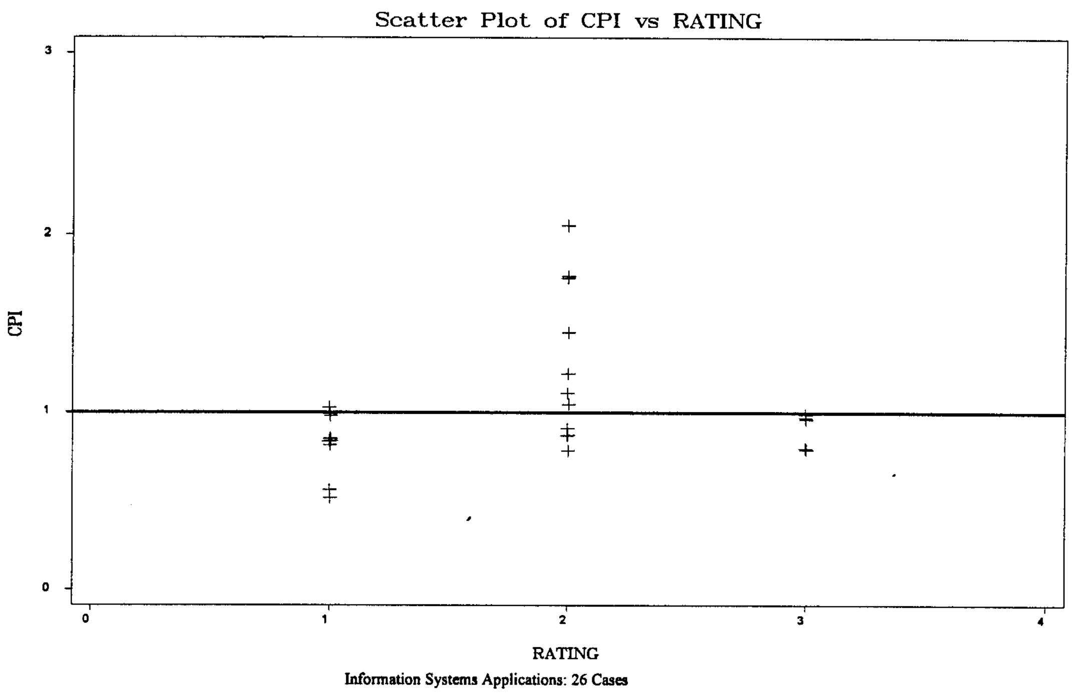

Figure I-3 Scatter Plot of CPI versus Rating for Information Systems Applications

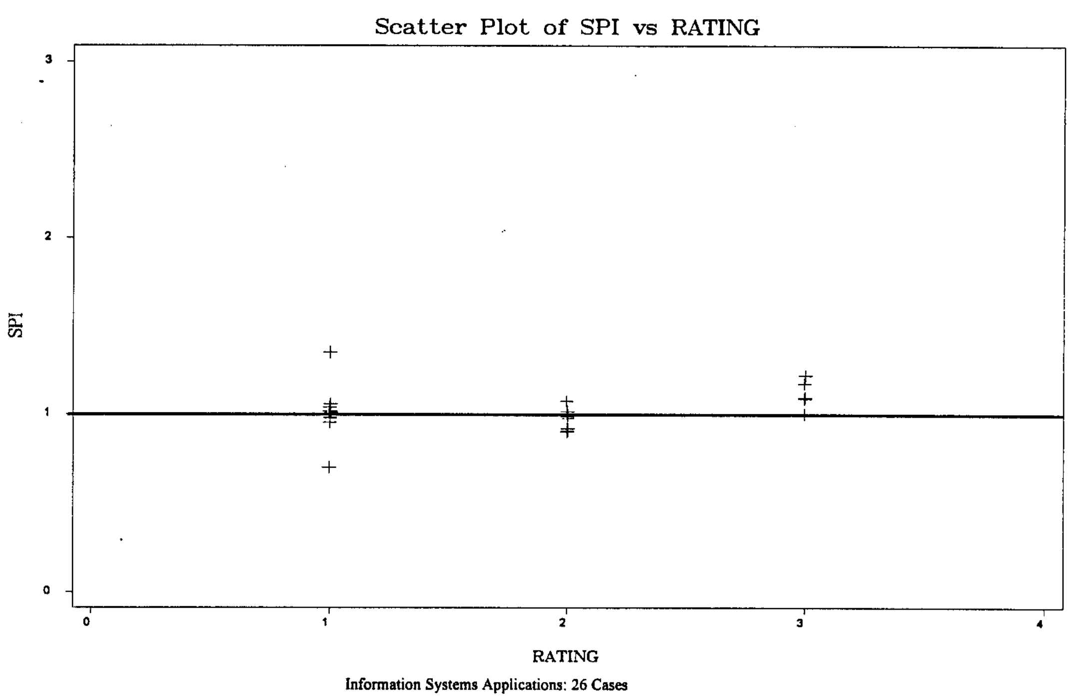

Figure I-4 Scatter Plot of SPI versus Rating for Information Systems Applications

2. *Box and Whiskers Plots of CPI and SPI*

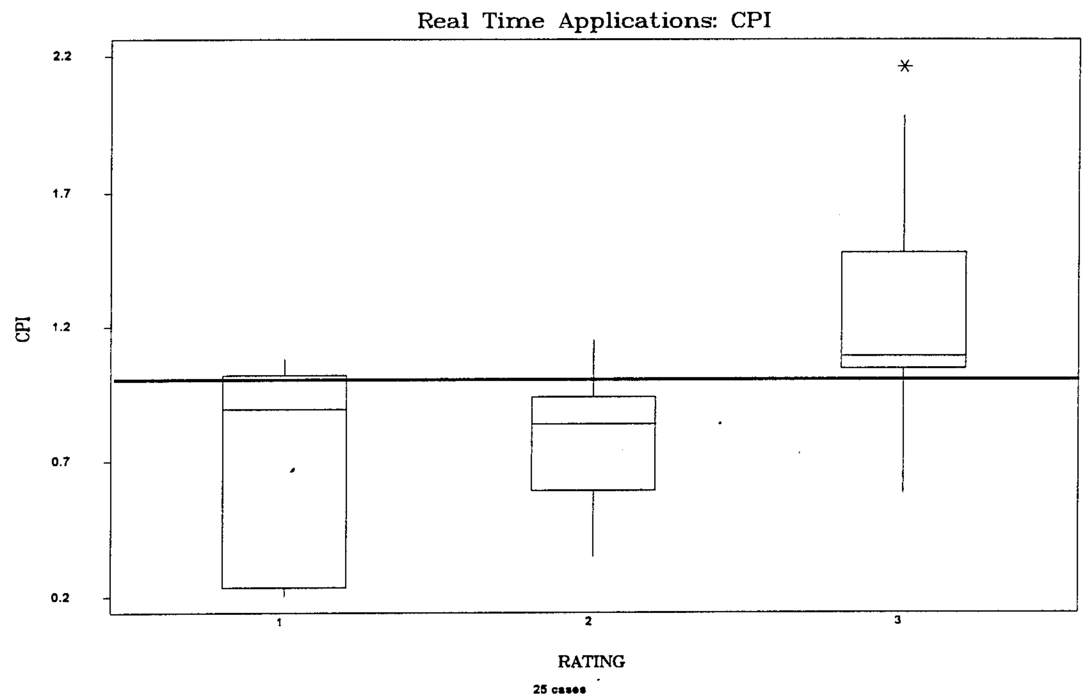

Figure I-5 Box and Whiskers Plot for CPI Real-Time Applications

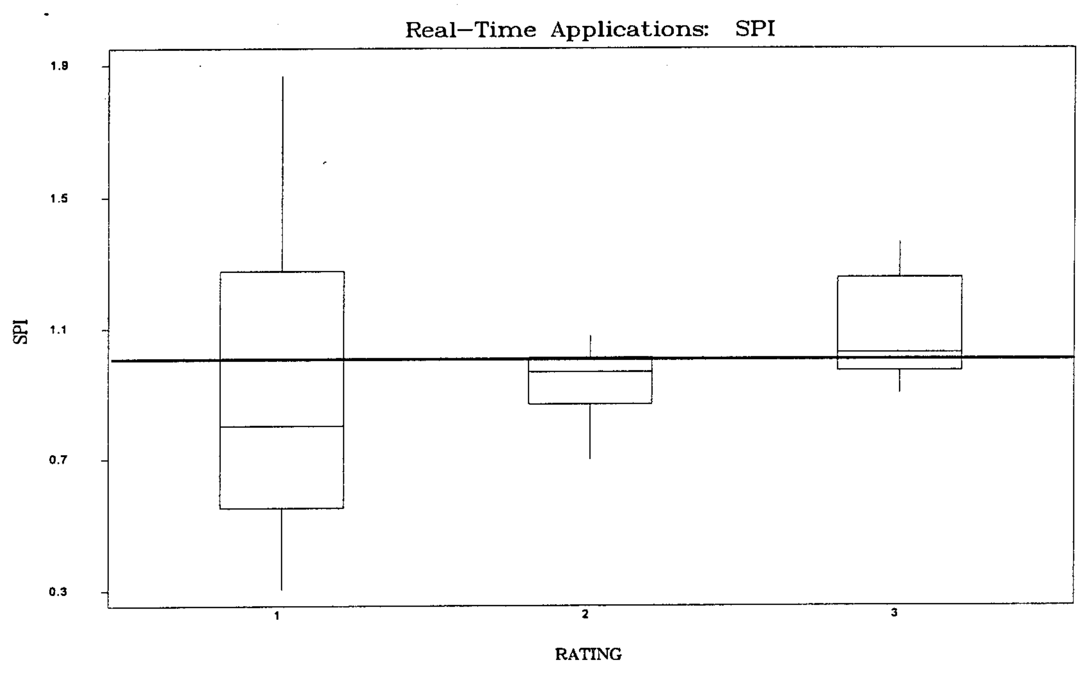

Figure I-6 Box and Whiskers Plot for SPI Real-Time Applications

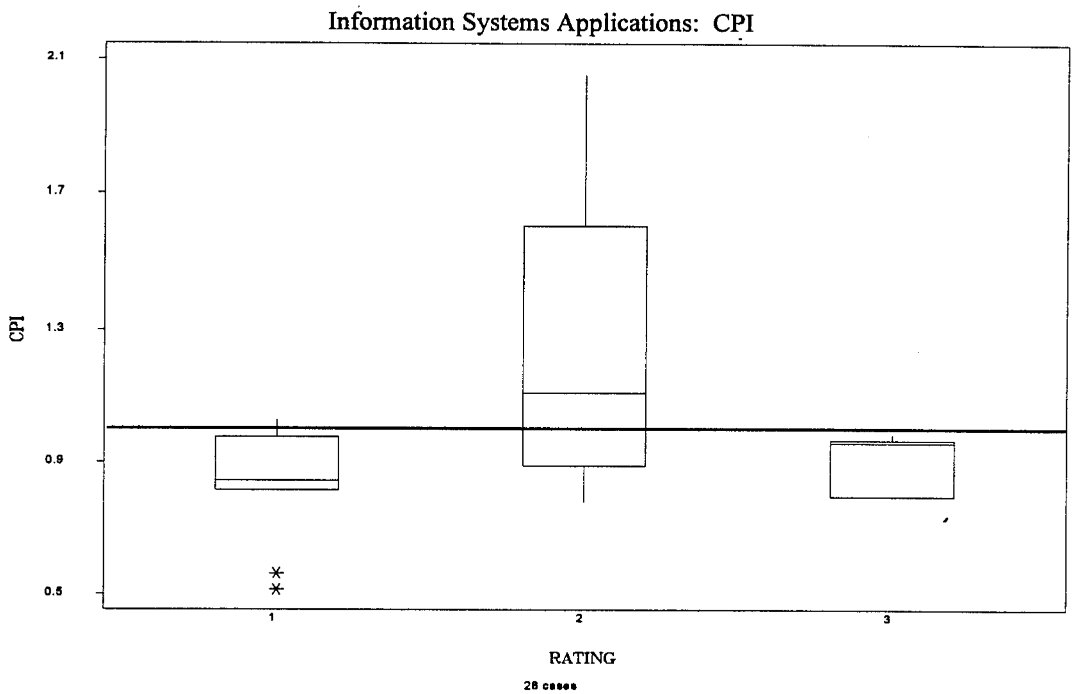

Figure I-7 Box and Whiskers Plot for CPI Information Systems Applications

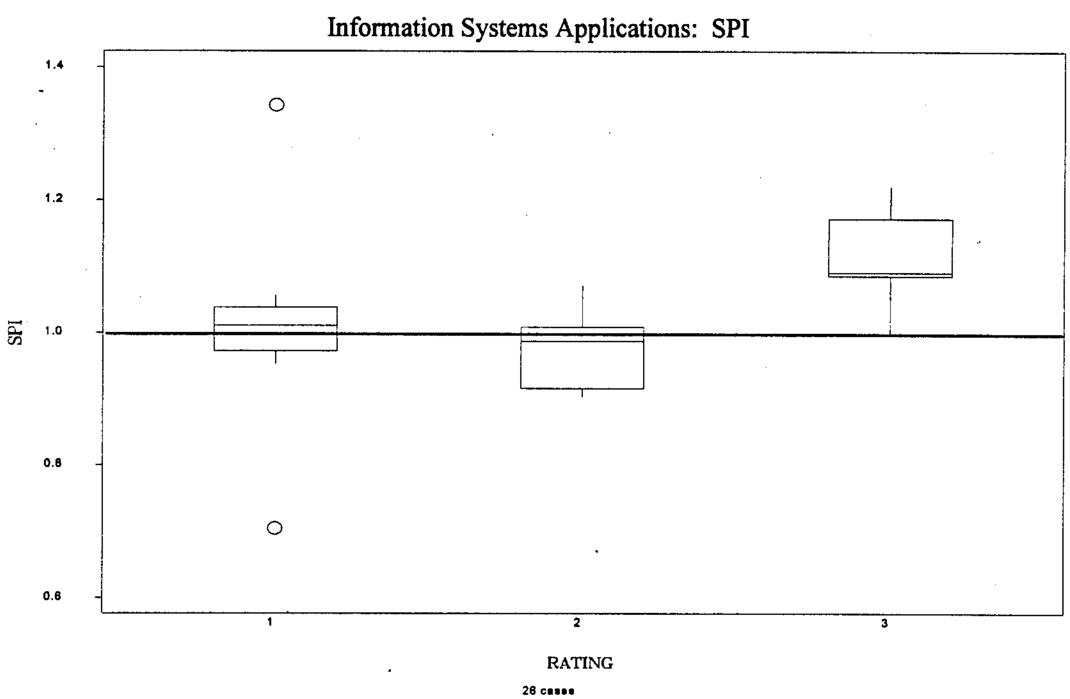

Figure I-8 Box and Whiskers Plot for SPI Information Systems Applications

3. Kruskal-Wallis Tests and Multiple Comparison Tests

Table I-1
Kruskal-Wallis for CPI for Real-Time Applications

```
KRUSKAL-WALLIS ONE-WAY NONPARAMETRIC AOV FOR CPI BY RATING

          MEAN   SAMPLE
 RATING   RANK   SIZE
      1    8.7      6
      2    8.9      7
      3   17.6     12
TOTAL     13.0     25

KRUSKAL-WALLIS STATISTIC                 8.9519
P-VALUE, USING CHI-SQUARED APPROXIMATION    0.0114
```

Table I-2
Multiple Comparison Matrix for CPI for Real-Time Applications

			Rating		
Rating	n	Mean Rank	1	2	3
1	6	8.7		--	--
2	7	8.9	-7.31		--
3	12	17.9	2.451	2.58	

K-W Statistic of 8.9519, P=0.0114

Table I-3
Kruskal-Wallis for SPI for Real-Time Applications

```
KRUSKAL-WALLIS ONE-WAY NONPARAMETRIC AOV FOR SPI BY RATING

          MEAN   SAMPLE
 RATING   RANK   SIZE
      1    9.8      6
      2   10.4      7
      3   16.1     12
TOTAL     13.0     25

KRUSKAL-WALLIS STATISTIC                 4.0714
P-VALUE, USING CHI-SQUARED APPROXIMATION    0.1306
```

Table I-4
Multiple Comparison Matrix for SPI for Real-Time Applications

			Rating		
Rating	n	Mean Rank	1	2	3
1	6	9.8		--	--
2	7	10.4	-6.91		--
3	12	16.1	-0.449	-0.72	

K-W Statistic of 4.0714, P=0.1306

Table I-5
Kruskal-Wallis for CPI for Information Systems Applications

```
KRUSKAL-WALLIS ONE-WAY NONPARAMETRIC AOV FOR CPI BY RATING

        MEAN   SAMPLE
RATING  RANK   SIZE
     1   9.5     10
     2  18.2     11
     3  11.2      5
TOTAL   13.5     26

KRUSKAL-WALLIS STATISTIC              7.3088
P-VALUE, USING CHI-SQUARED APPROXIMATION    0.0259
```

Table I-6
Multiple Comparison Matrix for CPI for Information Systems Applications

			Rating		
Rating	n	Mean Rank	1	2	3
1	10	9.5		--	--
2	11	18.2	2.571		--
3	5	11.2	-5.983	-0.566	

K-W Statistic of 7.3088, P=0.0259

Table I-7
Kruskal-Wallis for SPI for Information Systems Applications

```
KRUSKAL-WALLIS ONE-WAY NONPARAMETRIC AOV FOR SPI BY RATING

        MEAN   SAMPLE
RATING  RANK   SIZE
     1  13.3     10
     2  10.2     11
     3  21.2      5
TOTAL   13.5     26

KRUSKAL-WALLIS STATISTIC              7.1545
P-VALUE, USING CHI-SQUARED APPROXIMATION    0.0280
```

Table I-8
Multiple Comparison Matrix for SPI for Information Systems Applications

			Rating		
Rating	n	Mean Rank	1	2	3
1	10	13.3		--	--
2	11	10.2	-3.029		--
3	5	21.2	0.217	3.434	

K-W Statistic of 7.1545, P=0.0280

4. *Descriptive Statistics*

Table I-9
Descriptive Statistics for Real-Time Applications

	Rating=1 CPI	Rating=1 SPI	Rating=2 CPI	Rating=2 SPI	Rating=3 CPI	Rating=3 SPI
N	6	6	7	7	12	12
Mean	0.7202	0.9333	0.7705	0.9280	1.2592	1.1024
Std Dev	0.4024	0.5620	0.2966	0.1304	0.4556	0.1649
Min	0.2019	0.3028	0.3496	0.6978	0.5808	0.8998
Median	0.8928	0.8015	0.8368	0.9660	1.0890	1.0252
Max	1.0788	1.8676	1.1479	1.0774	2.1602	1.3652
MAD	0.1560	0.3625	0.1089	0.0814	0.0802	0.1085
Skew	-0.5187	0.6673	-0.4693	-0.6303	0.8288	0.3558

Table I-10
Descriptive Statistics for Information Systems Applications

	Rating=1 CPI	Rating=1 SPI	Rating=2 CPI	Rating=2 SPI	Rating=3 CPI	Rating=3 SPI
N	10	10	11	11	5	5
Mean	0.8250	1.0088	1.2581	0.9742	0.9003	1.1143
Std Dev	0.1688	0.1536	0.4360	0.0556	0.0972	0.0852
Min	0.5143	0.7044	0.7792	0.9043	0.7924	1.0000
Median	0.8437	1.0112	1.1114	0.9878	0.9599	1.0919
Max	1.0288	1.3428	2.0506	1.0718	0.9856	1.2207
MAD	0.0810	0.0336	0.2448	0.0321	0.0256	0.0805
Skew	-0.7858	0.2900	0.6107	0.0822	-0.3803	-0.0559

The Kruskal-Wallis H Test Statistic

$$H := \frac{12}{N \cdot (N+1)} \cdot \sum_{j=1}^{k} \frac{\left[R_j - \frac{n_j \cdot (N+1)}{2}\right]^2}{n_j}$$

Multiple Comparison Inequality

$$\left|\overline{R_i} - \overline{R_j}\right| \le z \cdot \sqrt{\frac{N \cdot (N+1)}{12} \cdot \left(\frac{1}{n_i} + \frac{1}{n_j}\right)}$$

where
k = number of samples
R_i, R_j = sum of the ranks in the ith and jth sample, respectively
n_i, n_j = number of observations in the ith and jth sample, respectively
N = total number of observations

Mean Rank of Ratings 1, 2, & 3 for CPI
$R_{cpi} := (8.7\ \ 8.9\ \ 17.9)$

Mean Rank of Ratings 1, 2, & 3 for SPI
$R_{spi} := (9.8\ \ 10.4\ \ 16.1)$

Number of Observations in Ratings 1, 2, & 3
$n := (6\ \ 7\ \ 12)$

Calculation of N $\quad N := n_{1,1} + n_{1,2} + n_{1,3} \quad N = 25$

Critical z Value - Corresponds to a .2 level of significance and provides at least a 80% level of confidence.
$z := 1.834$

<u>Multiple Comparisons</u> - "which populations differ from which others"
- If the difference in mean rank (DMR) between 2 samples is greater than the right-hand side (RHS) of the multiple comparison inequality above, there is a statistically significant difference between the 2 samples.
- The direction of the significant difference is determined by noting for each pair which sample has the larger sample mean.

Comparision of Rating 1 and 2

$DMR_{cpi} := \left|R_{cpi_{1,1}} - R_{cpi_{1,2}}\right| \quad DMR_{cpi} = 0.2$

$DMR_{spi} := \left|R_{spi_{1,1}} - R_{spi_{1,2}}\right| \quad DMR_{spi} = 0.6$

$$RHS := z \cdot \sqrt{\frac{N \cdot (N+1)}{12} \cdot \left(\frac{1}{n_{1,1}} + \frac{1}{n_{1,2}}\right)} \quad RHS = 7.51$$

For CPI $DMR_{cpi} - RHS = -7.31$ — no significant difference <==
For SPI $DMR_{spi} - RHS = -6.91$ — no significant difference <==

Comparision of Rating 2 and 3

$DMR_{cpi} := \left|R_{cpi_{1,2}} - R_{cpi_{1,3}}\right| \quad DMR_{cpi} = 9$

$DMR_{spi} := \left|R_{spi_{1,2}} - R_{spi_{1,3}}\right| \quad DMR_{spi} = 5.7$

$$RHS := z \cdot \sqrt{\frac{N \cdot (N+1)}{12} \cdot \left(\frac{1}{n_{1,2}} + \frac{1}{n_{1,3}}\right)} \quad RHS = 6.42$$

For CPI $DMR_{cpi} - RHS = 2.58$ — CPI at Rating 3 > CPI at Rating 2 <==
For SPI $DMR_{spi} - RHS = -0.72$ — no significant difference <==

Comparision of Rating 1 and 3

$DMR_{cpi} := \left|R_{cpi_{1,1}} - R_{cpi_{1,3}}\right| \quad DMR_{cpi} = 9.2$

$DMR_{spi} := \left|R_{spi_{1,1}} - R_{spi_{1,3}}\right| \quad DMR_{spi} = 6.3$

$$RHS := z \cdot \sqrt{\frac{N \cdot (N+1)}{12} \cdot \left(\frac{1}{n_{1,1}} + \frac{1}{n_{1,3}}\right)} \quad RHS = 6.749$$

For CPI $DMR_{cpi} - RHS = 2.451$ — CPI at Rating 3 > CPI at Rating 1 <==
For SPI $DMR_{spi} - RHS = -0.449$ — no significant difference <==

Figure I-9 Calculations for Multiple Comparison Test for Real-Time Applications

6. Multiple Comparison Calculations for Information Systems Applications

The Kruskal-Wallis H Test Statistic

$$H := \frac{12}{N \cdot (N+1)} \cdot \sum_{j=1}^{k} \frac{\left[R_j - \frac{n_j \cdot (N+1)}{2}\right]^2}{n_j}$$

Multiple Comparison Inequality

$$\left|\overline{R_i} - \overline{R_j}\right| \le z \cdot \sqrt{\frac{N \cdot (N+1)}{12} \cdot \left(\frac{1}{n_i} + \frac{1}{n_j}\right)}$$

where
k = number of samples
R_i, R_j = sum of the ranks in the ith and jth sample, respectively
n_i, n_j = number of observations in the ith and jth sample, respectively
N = total number of observations

Mean Rank of Ratings 1, 2, & 3 for CPI
$R_{cpi} := (9.5\ \ 18.2\ \ 11.2)$

Mean Rank of Ratings 1, 2, & 3 for SPI
$R_{spi} := (13.3\ \ 10.2\ \ 21.2)$

Number of Observations in Ratings 1, 2, & 3
$n := (10\ \ 11\ \ 5)$

Calculation of N $\quad N := n_{1,1} + n_{1,2} + n_{1,3} \quad N = 26$

Critical z Value - Corresponds to a .2 level of significance and provides at least a 80% level of confidence.
$z := 1.834$

__Multiple Comparisons__ - "which populations differ from which others"
- If the difference in mean rank (DMR) between 2 samples is greater than the right-hand side (RHS) of the multiple comparison inequality above, there is a statistically significant difference between the 2 samples.
- The direction of the significant difference is determined by noting for each pair which sample has the larger sample mean.

Comparision of Rating 1 and 2

$DMR_{cpi} := \left|R_{cpi_{1,1}} - R_{cpi_{1,2}}\right| \quad DMR_{cpi} = 8.7$

$DMR_{spi} := \left|R_{spi_{1,1}} - R_{spi_{1,2}}\right| \quad DMR_{spi} = 3.1$

$RHS := z \cdot \sqrt{\frac{N \cdot (N+1)}{12} \cdot \left(\frac{1}{n_{1,1}} + \frac{1}{n_{1,2}}\right)} \quad RHS = 6.129$

For CPI $\ DMR_{cpi} - RHS = 2.571$ — CPI at Rating 2 > CPI at Rating 1 <===
For SPI $\ DMR_{spi} - RHS = -3.029$ — no significant difference <===

Comparision of Rating 2 and 3

$DMR_{cpi} := \left|R_{cpi_{1,2}} - R_{cpi_{1,3}}\right| \quad DMR_{cpi} = 7$

$DMR_{spi} := \left|R_{spi_{1,2}} - R_{spi_{1,3}}\right| \quad DMR_{spi} = 11$

$RHS := z \cdot \sqrt{\frac{N \cdot (N+1)}{12} \cdot \left(\frac{1}{n_{1,2}} + \frac{1}{n_{1,3}}\right)} \quad RHS = 7.566$

For CPI $\ DMR_{cpi} - RHS = -0.566$ — no significant difference <===
For SPI $\ DMR_{spi} - RHS = 3.434$ — SPI at Rating 3 > SPI at Rating 2 <===

Comparision of Rating 1 and 3

$DMR_{cpi} := \left|R_{cpi_{1,1}} - R_{cpi_{1,3}}\right| \quad DMR_{cpi} = 1.7$

$DMR_{spi} := \left|R_{spi_{1,1}} - R_{spi_{1,3}}\right| \quad DMR_{spi} = 7.9$

$RHS := z \cdot \sqrt{\frac{N \cdot (N+1)}{12} \cdot \left(\frac{1}{n_{1,1}} + \frac{1}{n_{1,3}}\right)} \quad RHS = 7.683$

For CPI $\ DMR_{cpi} - RHS = -5.983$ — no significant difference <===
For SPI $\ DMR_{spi} - RHS = 0.217$ — SPI at Rating 3 > SPI at Rating 1 <===

Figure I-10 Calculations for Multiple Comparison Test for Information Systems Applications

Appendix J: Support for Analysis of Moderator "Language"

This appendix contains the complete set of plots, tables, and calculations supporting section 5.2.2. The plots and the Kruskal-Wallis tables for nonparametric analysis of variance are from the Statistix 4.0 computer program. The multiple comparison calculations were performed using Mathcad 4.0. Note that we did not abridge the data as we transcribed it from these computer programs into the report, and thus the number of digits reported in each calculation are not necessarily significant.

Ada, as the official "standard" higher order language (HOL) of the DoD, is mandated for all new software development programs. This requirement to use Ada may impose difficulties on software development contractors if they have little experience with Ada, or if Ada is not their preferred language. On the other hand Ada is a powerful language which imposes rigorous discipline in the development process, and thus may provide benefits in the testing and integration phases of development. Thus it is important to determine if such a significant program characteristic has any effect on the correlation between rating and performance.

1. Scatter Plots of CPI and SPI

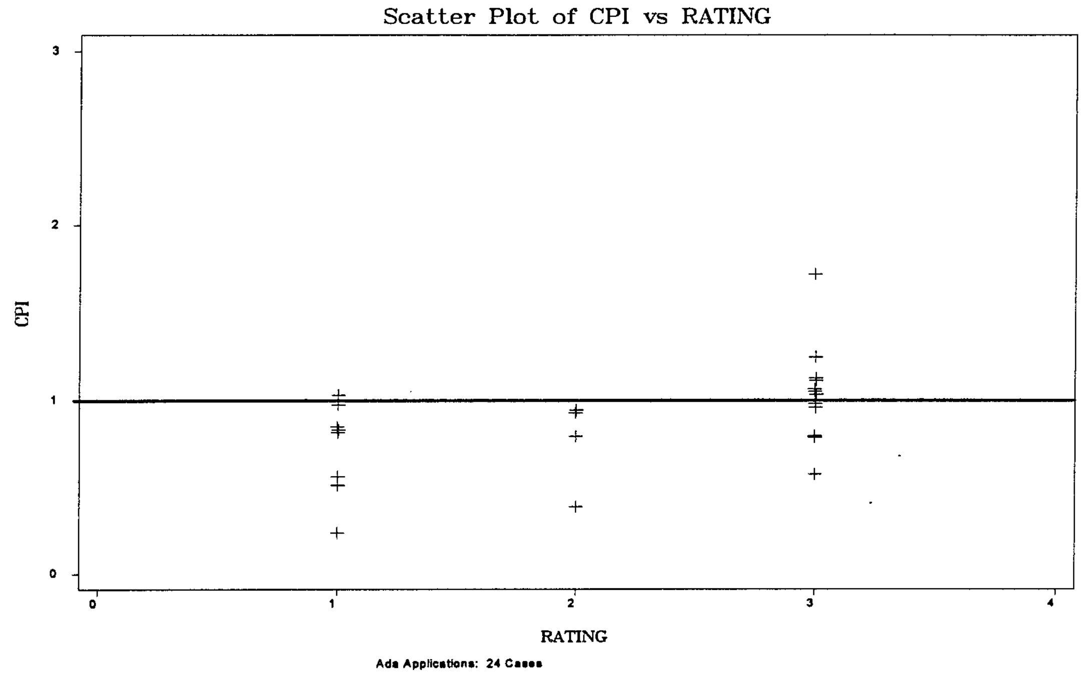

Figure J-1 Scatter Plot of CPI versus Rating for Ada Applications

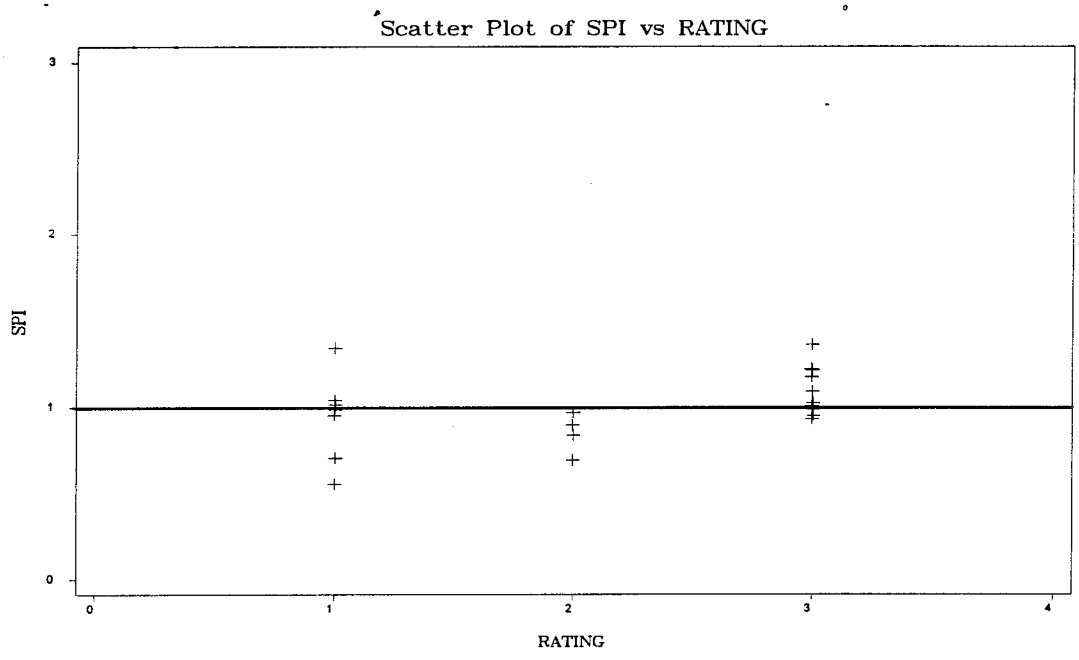

Figure J-2 Scatter Plot of SPI versus Rating for Ada Applications

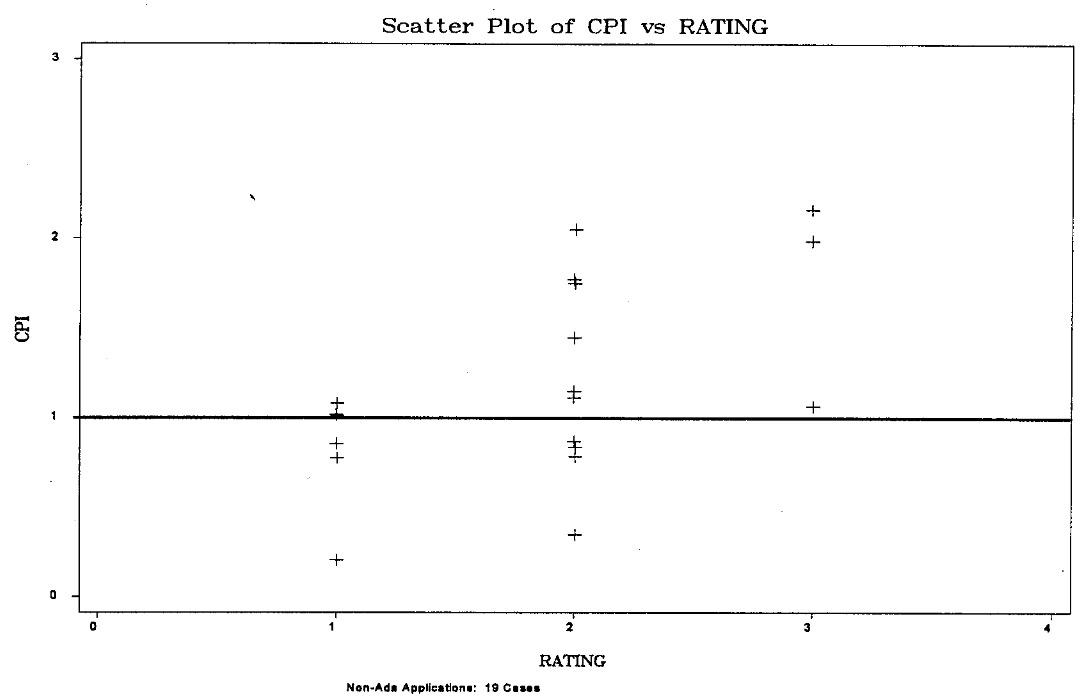

Figure J-3 Scatter Plot of CPI versus Rating for Non-Ada Applications

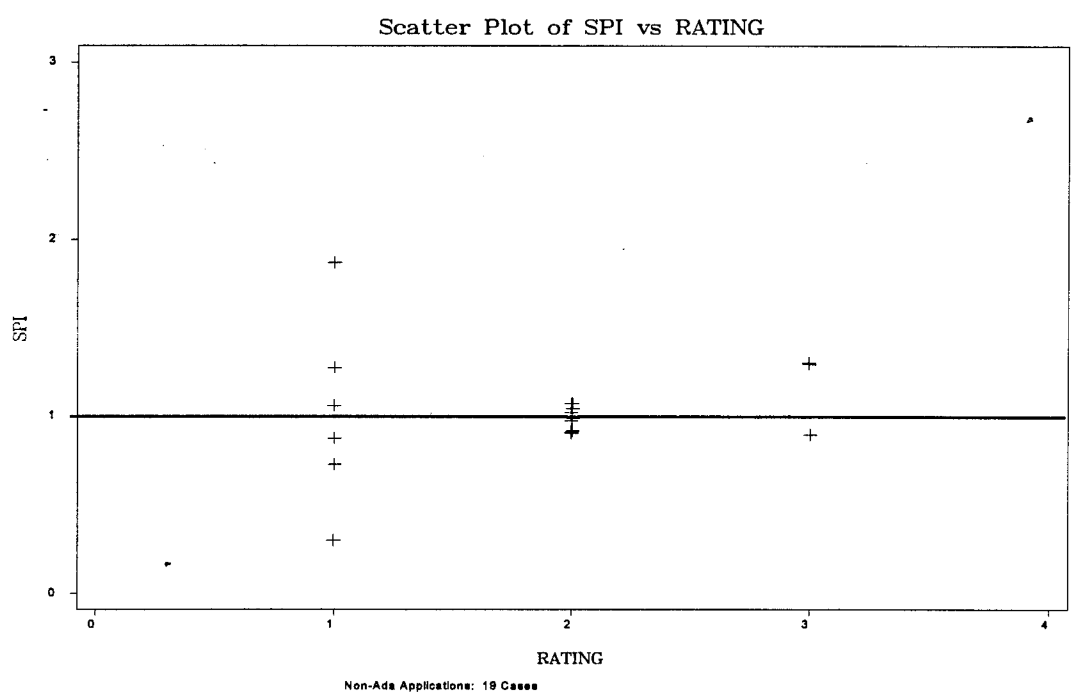

Figure J-4 Scatter Plot of SPI versus Rating for Non-Ada Applications

2. Box and Whisker Plots of CPI and SPI

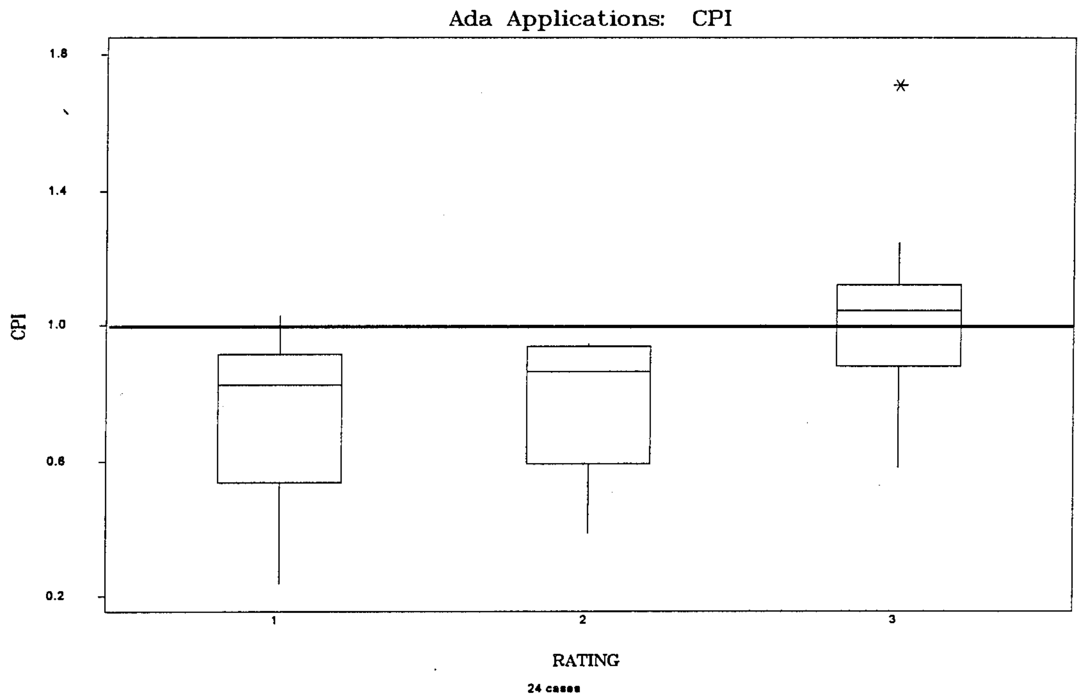

Figure J-5 Box and Whiskers Plot for CPI for Ada Applications

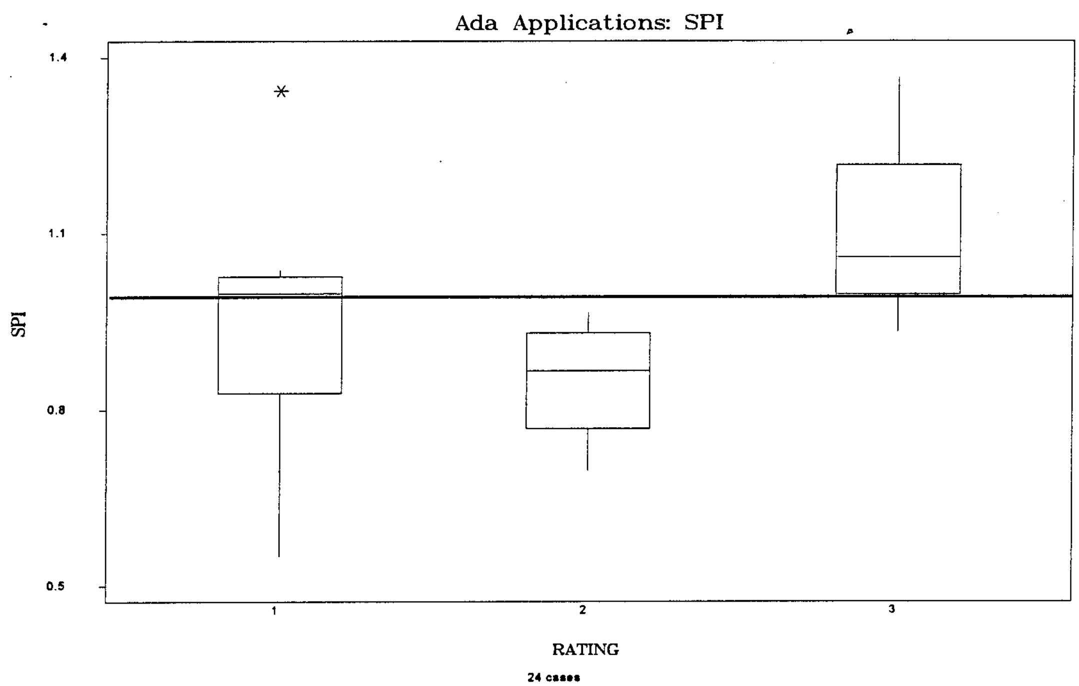

Figure J-6 Box and Whiskers Plot for SPI for Ada Applications

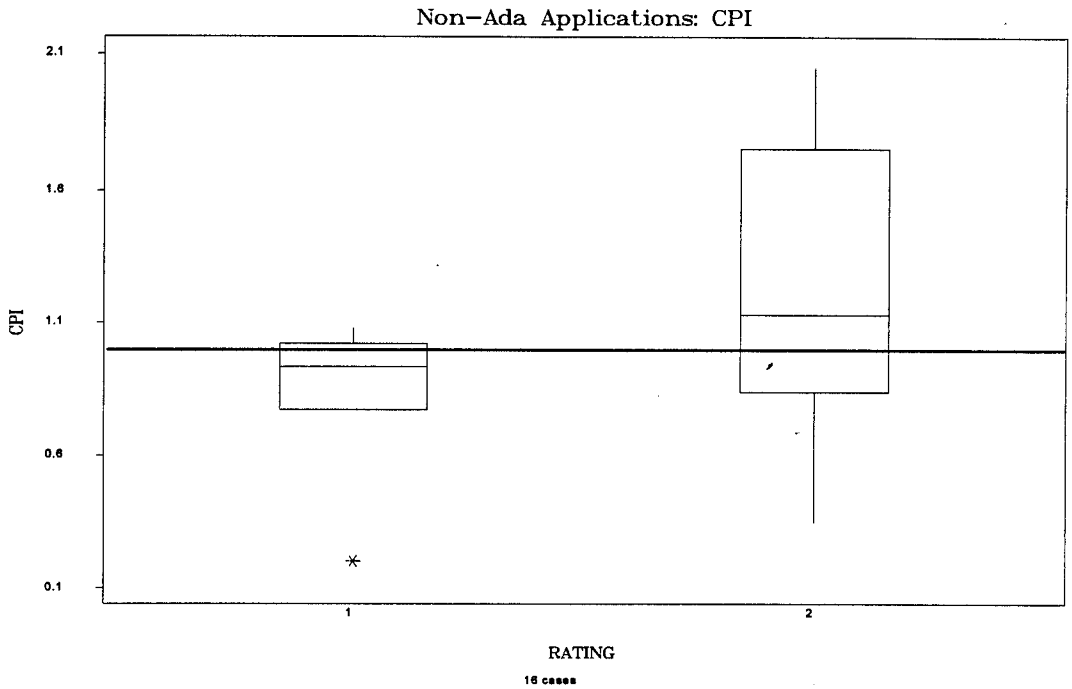

Figure J-7 Box and Whiskers Plot for CPI for Non-Ada Applications

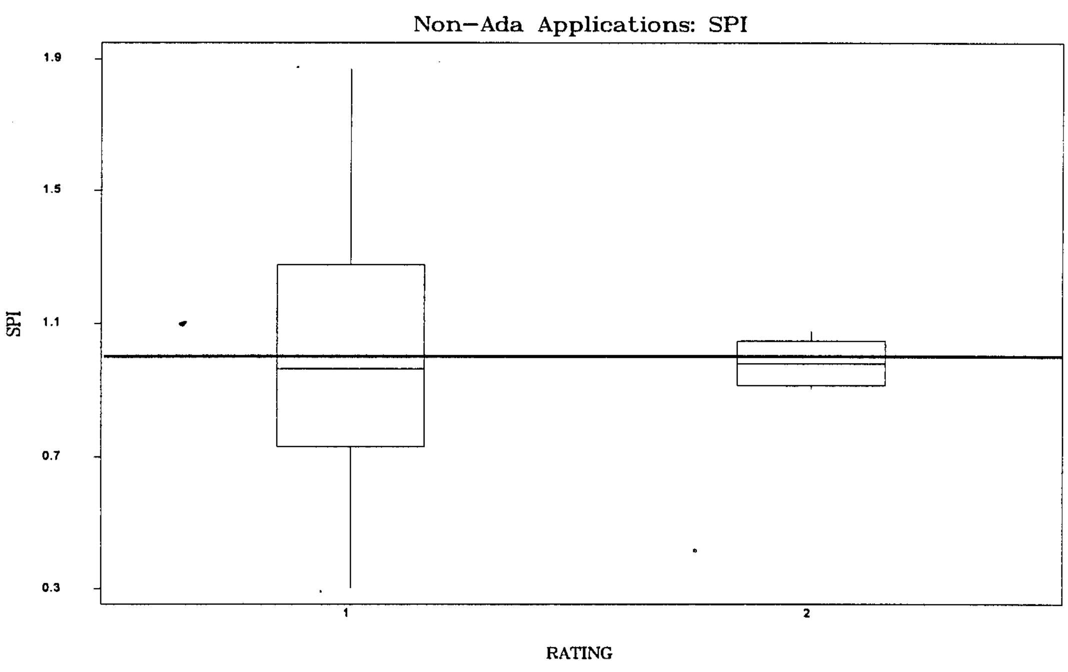

Figure J-8 Box and Whiskers Plot for SPI for Non-Ada Applications

3. Kruskal-Wallis and Multiple Comparison Tests

Table J-1
Kruskal-Wallis for CPI for Ada Applications

KRUSKAL-WALLIS ONE-WAY NONPARAMETRIC AOV FOR CPI BY RATING

RATING	MEAN RANK	SAMPLE SIZE
1	8.8	8
2	8.8	4
3	16.3	12
TOTAL	12.5	24

KRUSKAL-WALLIS STATISTIC 6.7500
P-VALUE, USING CHI-SQUARED APPROXIMATION 0.0342

Table J-2
Multiple Comparison Matrix for CPI for Ada Applications

			Rating		
Rating	n	Mean Rank	1	2	3
1	8	8.8		--	--
2	4	8.8	-7.941		--
3	12	16.3	1.581	0.013	

K-W Statistic of 6.7500, P=0.0342

Table J-3
Kruskal-Wallis for SPI for Ada Applications

KRUSKAL-WALLIS ONE-WAY NONPARAMETRIC AOV FOR SPI BY RATING

RATING	MEAN RANK	SAMPLE SIZE
1	11.0	8
2	5.0	4
3	16.0	12
TOTAL	12.5	24

KRUSKAL-WALLIS STATISTIC 7.8000
P-VALUE, USING CHI-SQUARED APPROXIMATION 0.0202

Table J-4
Multiple Comparison Matrix for SPI for Ada Applications

			Rating		
Rating	n	Mean Rank	1	2	3
1	8	11.0		--	--
2	4	5.0	-1.941		--
3	12	16.0	-0.919	3.513	

K-W Statistic of 7.800, P=0.0202

Table J-5
Kruskal-Wallis for CPI for Non-Ada Applications

KRUSKAL-WALLIS ONE-WAY NONPARAMETRIC AOV FOR CPI BY RATING

RATING	MEAN RANK	SAMPLE SIZE
1	6.3	6
2	10.6	10
3	15.3	3
TOTAL	10.0	19

KRUSKAL-WALLIS STATISTIC 5.3558
P-VALUE, USING CHI-SQUARED APPROXIMATION 0.0687

Table J-6
Multiple Comparison Matrix for CPI for Non-Ada Applications

			Rating		
Rating	n	Mean Rank	1	2	3
1	6	6.3		--	--
2	10	10.6	-1.029		--
3	3	15.3	1.702	-2.094	

K-W Statistic of 5.3558, P=0.0687

Table J-7
Kruskal-Wallis for SPI for Non-Ada Applications

KRUSKAL-WALLIS ONE-WAY NONPARAMETRIC AOV FOR SPI BY RATING

RATING	MEAN RANK	SAMPLE SIZE
1	9.0	6
2	9.7	10
3	13.0	3
TOTAL	10.0	19

KRUSKAL-WALLIS STATISTIC 1.0705
P-VALUE, USING CHI-SQUARED APPROXIMATION 0.5855

Table J-8
Multiple Comparison Matrix for SPI for Non-Ada Applications

			Rating		
Rating	n	Mean Rank	1	2	3
1	6	9.0		--	--
2	10	9.7	-4.629		--
3	3	13.0	-3.298	-3.494	

K-W Statistic of 1.0705, P=0.5855

4. *Descriptive Statistics*

Table J-9
Descriptive Statistics for Ada Applications

	Rating=1 CPI	Rating=1 SPI	Rating=2 CPI	Rating=2 SPI	Rating=3 CPI	Rating=3 SPI
N	8	8	4	4	12	12
Mean	0.7270	0.9501	0.7648	0.8496	1.0375	1.1012
Std Dev	0.2675	0.2364	0.2606	0.1139	0.2771	0.1345
Min	0.2363	0.5507	0.3863	0.6978	0.5808	0.9335
Median	0.8242	0.9981	0.8637	0.8673	1.0425	1.0590
Max	1.0288	1.3428	0.9457	0.9660	1.7097	1.3652
MAD	0.1787	0.0424	0.0729	0.0639	0.0836	0.1093
Skew	-0.6825	-0.1938	-0.9573	-0.4744	0.8438	0.4835

Table J-10
Descriptive Statistics for Non-Ada Applications

	Rating=1 CPI	Rating=1 SPI	Rating=2 CPI	Rating=2 SPI	Rating=3 CPI	Rating=3 SPI
N	6	6	10	10	3	3
Mean	0.8224	1.0175	1.2126	0.9823	1.7365	1.1664
Std Dev	0.3254	0.5301	0.5353	0.0692	0.5862	0.2310
Min	0.2019	0.3028	0.3496	0.9043	1.0675	0.8998
Median	0.9321	0.9643	1.1296	0.9803	1.9819	1.2920
Max	1.0788	1.8676	2.0506	1.0774	2.1602	1.3073
MAD	0.1166	0.2727	0.3377	0.0662	0.1783	0.0153
Skew	-1.3198	0.3527	0.0912	0.1613	-0.6343	-0.7036

5. Multiple Comparison Calculations for Ada Applications

The Kruskal-Wallis H Test Statistic

$$H := \frac{12}{N\cdot(N+1)}\cdot\sum_{j=1}^{k}\frac{\left[R_j - \frac{n_j\cdot(N+1)}{2}\right]^2}{n_j}$$

Multiple Comparison Inequality

$$\left|\overline{R_i} - \overline{R_j}\right| \le z\cdot\sqrt{\frac{N\cdot(N+1)}{12}\cdot\left(\frac{1}{n_i}+\frac{1}{n_j}\right)}$$

where

k = number of samples

R_i, R_j = sum of the ranks in the ith and jth sample, respectively

n_i, n_j = number of observations in the ith and jth sample, respectively

N = total number of observations

Mean Rank of Ratings 1, 2, & 3 for CPI

$R_{cpi} := (8.8\ \ 8.8\ \ 16.3)$

Mean Rank of Ratings 1, 2, & 3 for SPI

$R_{spi} := (11.0\ \ 5.0\ \ 16.0)$

Number of Observations in Ratings 1, 2, & 3

$n := (8\ \ 4\ \ 12)$

Calculation of N $\quad N := n_{1,1} + n_{1,2} + n_{1,3} \quad N = 24$

Critical z Value - Corresponds to a .2 level of significance and provides at least a 80% level of confidence.

$z := 1.834$

Multiple Comparisons - "which populations differ from which others"

- If the difference in mean rank (DMR) between 2 samples is greater than the right-hand side (RHS) of the multiple comparison inequality above, there is a statistically significant difference between the 2 samples.
- The direction of the significant difference is determined by noting for each pair which sample has the larger sample mean.

Comparision of Rating 1 and 2

$DMR_{cpi} := \left|R_{cpi_{1,1}} - R_{cpi_{1,2}}\right| \quad DMR_{cpi} = 0$

$DMR_{spi} := \left|R_{spi_{1,1}} - R_{spi_{1,2}}\right| \quad DMR_{spi} = 6$

$RHS := z\cdot\sqrt{\frac{N\cdot(N+1)}{12}\cdot\left(\frac{1}{n_{1,1}}+\frac{1}{n_{1,2}}\right)} \quad RHS = 7.941$

For CPI $DMR_{cpi} - RHS = -7.941$ — no significant difference <==

For SPI $DMR_{spi} - RHS = -1.941$ — no significant difference <==

Comparision of Rating 2 and 3

$DMR_{cpi} := \left|R_{cpi_{1,2}} - R_{cpi_{1,3}}\right| \quad DMR_{cpi} = 7.5$

$DMR_{spi} := \left|R_{spi_{1,2}} - R_{spi_{1,3}}\right| \quad DMR_{spi} = 11$

$RHS := z\cdot\sqrt{\frac{N\cdot(N+1)}{12}\cdot\left(\frac{1}{n_{1,2}}+\frac{1}{n_{1,3}}\right)} \quad RHS = 7.487$

For CPI $DMR_{cpi} - RHS = 0.013$ — CPI at Rating 3 > CPI at Rating 2 <==

For SPI $DMR_{spi} - RHS = 3.513$ — SPI at Rating 3 > SPI at Rating 2 <==

Comparision of Rating 1 and 3

$DMR_{cpi} := \left|R_{cpi_{1,1}} - R_{cpi_{1,3}}\right| \quad DMR_{cpi} = 7.5$

$DMR_{spi} := \left|R_{spi_{1,1}} - R_{spi_{1,3}}\right| \quad DMR_{spi} = 5$

$RHS := z\cdot\sqrt{\frac{N\cdot(N+1)}{12}\cdot\left(\frac{1}{n_{1,1}}+\frac{1}{n_{1,3}}\right)} \quad RHS = 5.919$

For CPI $DMR_{cpi} - RHS = 1.581$ — CPI at Rating 3 > CPI at Rating 1 <==

For SPI $DMR_{spi} - RHS = -0.919$ — no significant difference <==

Figure J-9 Calculations for Multiple Comparison Test for Ada Applications

6. Multiple Comparison Calculations for Non-Ada Applications

The Kruskal-Wallis H Test Statistic

$$H := \frac{12}{N\cdot(N+1)} \cdot \sum_{j=1}^{k} \frac{\left[R_j - \frac{n_j\cdot(N+1)}{2}\right]^2}{n_j}$$

Multiple Comparison Inequality

$$\left|\overline{R_i} - \overline{R_j}\right| \le z\cdot\sqrt{\frac{N\cdot(N+1)}{12}\cdot\left(\frac{1}{n_i}+\frac{1}{n_j}\right)}$$

where
k = number of samples
R_i, R_j = sum of the ranks in the ith and jth sample, respectively
n_i, n_j = number of observations in the ith and jth sample, respectively
N = total number of observations

Mean Rank of Ratings 1, 2, & 3 for CPI	**Mean Rank of Ratings 1, 2, & 3 for SPI**	**Number of Observations in Ratings 1, 2, & 3**
$R_{cpi} := (6.3\ \ 10.6\ \ 15.3)$	$R_{spi} := (9.0\ \ 9.7\ \ 13.0)$	$n := (6\ \ 10\ \ 3)$

Calculation of N $\quad N := n_{1,1} + n_{1,2} + n_{1,3} \quad N = 19$

Critical z Value - Corresponds to a .2 level of significance and provides at least a 80% level of confidence.
$z := 1.834$

<u>**Multiple Comparisons**</u> - "which populations differ from which others"
- If the difference in mean rank (DMR) between 2 samples is greater than the right-hand side (RHS) of the multiple comparison inequality above, there is a statistically significant difference between the 2 samples.
- The direction of the significant difference is determined by noting for each pair which sample has the larger sample mean.

Comparision of Rating 1 and 2

$DMR_{cpi} := \left|R_{cpi_{1,1}} - R_{cpi_{1,2}}\right| \quad DMR_{cpi} = 4.3$

$DMR_{spi} := \left|R_{spi_{1,1}} - R_{spi_{1,2}}\right| \quad DMR_{spi} = 0.7$

$RHS := z\cdot\sqrt{\frac{N\cdot(N+1)}{12}\cdot\left(\frac{1}{n_{1,1}}+\frac{1}{n_{1,2}}\right)} \quad RHS = 5.329$

For CPI $DMR_{cpi} - RHS = -1.029$ — no significant difference <==
For SPI $DMR_{spi} - RHS = -4.629$ — no significant difference <==

Comparision of Rating 2 and 3

$DMR_{cpi} := \left|R_{cpi_{1,2}} - R_{cpi_{1,3}}\right| \quad DMR_{cpi} = 4.7$

$DMR_{spi} := \left|R_{spi_{1,2}} - R_{spi_{1,3}}\right| \quad DMR_{spi} = 3.3$

$RHS := z\cdot\sqrt{\frac{N\cdot(N+1)}{12}\cdot\left(\frac{1}{n_{1,2}}+\frac{1}{n_{1,3}}\right)} \quad RHS = 6.794$

For CPI $DMR_{cpi} - RHS = -2.094$ — CPI at Rating 3 > CPI at Rating 2 <==
For SPI $DMR_{spi} - RHS = -3.494$ — SPI at Rating 3 > SPI at Rating 2 <==

Comparision of Rating 1 and 3

$DMR_{cpi} := \left|R_{cpi_{1,1}} - R_{cpi_{1,3}}\right| \quad DMR_{cpi} = 9$

$DMR_{spi} := \left|R_{spi_{1,1}} - R_{spi_{1,3}}\right| \quad DMR_{spi} = 4$

$RHS := z\cdot\sqrt{\frac{N\cdot(N+1)}{12}\cdot\left(\frac{1}{n_{1,1}}+\frac{1}{n_{1,3}}\right)} \quad RHS = 7.298$

For CPI $DMR_{cpi} - RHS = 1.702$ — CPI at Rating 3 > CPI at Rating 1 <==
For SPI $DMR_{spi} - RHS = -3.298$ — no significant difference <==

Figure J-10 Calculations for Multiple Comparison Test for Non-Ada Applications

Appendix K: Data Supporting the Analysis of the Moderator "Project Size"

This appendix contains the complete set of plots, tables, and calculations supporting section 5.2.2. The plots and the Kruskal-Wallis tables for nonparametric analysis of variance are from the Statistix 4.0 computer program. The multiple comparison calculations were performed using Mathcad 4.0. Note that we did not abridge the data as we transcribed it from these computer programs into the report, and thus the number of digits reported in each calculation are not necessarily significant.

Project size is the key driver in nearly all software cost estimation models, including REVIC (Revised Enhanced Version of Intermediate COCOMO), SEER-Software Estimation Model, and PRICE-S. Thus, project size is a necessary moderator to evaluate, in terms of its effect upon the rating/performance correlation. Given the lack of uniformity in the definition of software project size (we gathered data in the form of KSLOC, DSI, Equivalent DSI, and DSI converted from bytes), we can at best only give approximate size distinctions. Thus, we chose to stratify our sample on the relatively common size categories: "Greater than 100K LOC" and "Less than 100K LOC." This level of distinction is fairly common in the literature when distinguishing between relatively large programs and relatively small programs. As stated above, with the questionable consistency of our size data, any finer distinction would be misleading. Projects which have no size associated with them, such as management or testing WBSs, were excluded from the analysis.

1. Scatter Plots of CPI and SPI

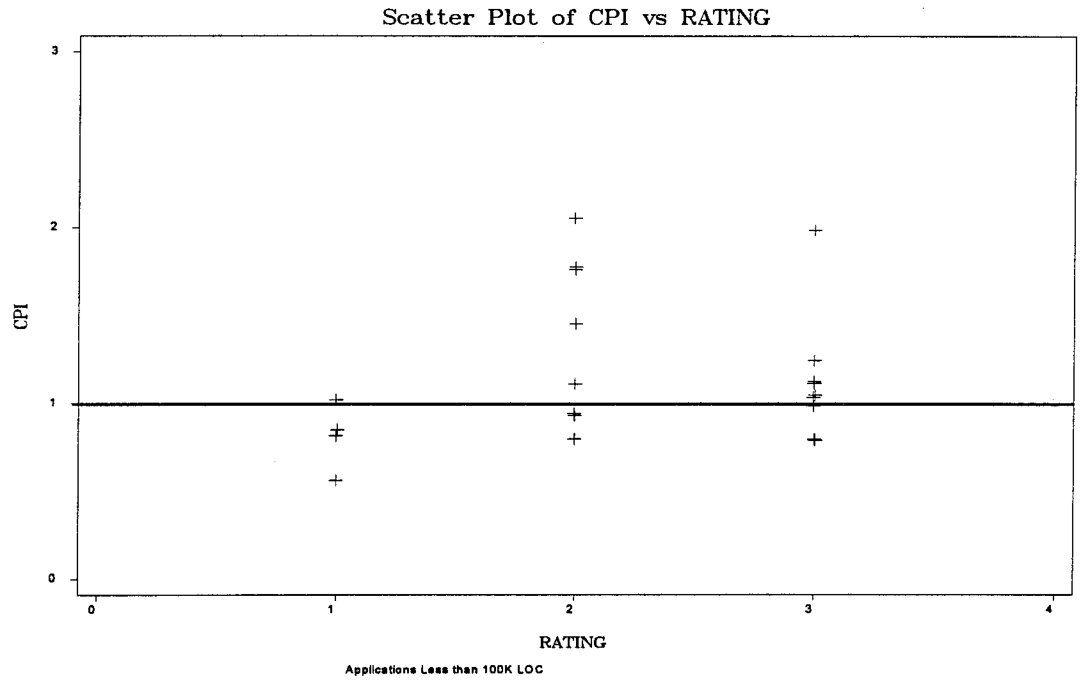

Figure K-1 Scatter Plot of CPI versus Rating for Applications Less than 100K (LOC)

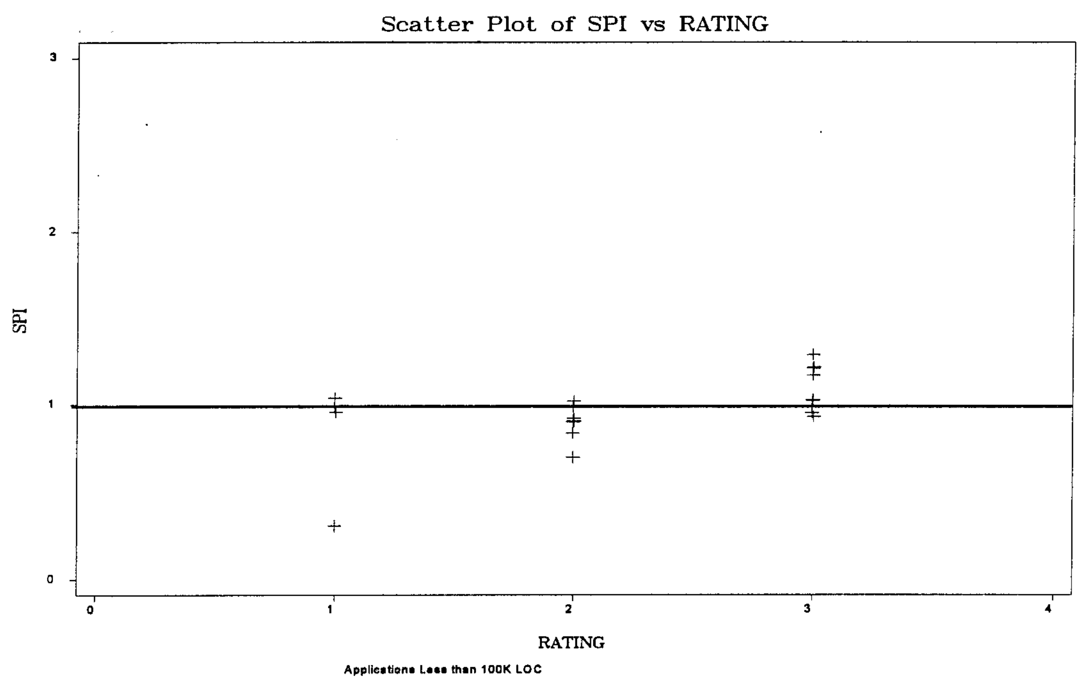

Figure K-2 Scatter Plot of SPI versus Rating for Applications Less than 100K (LOC)

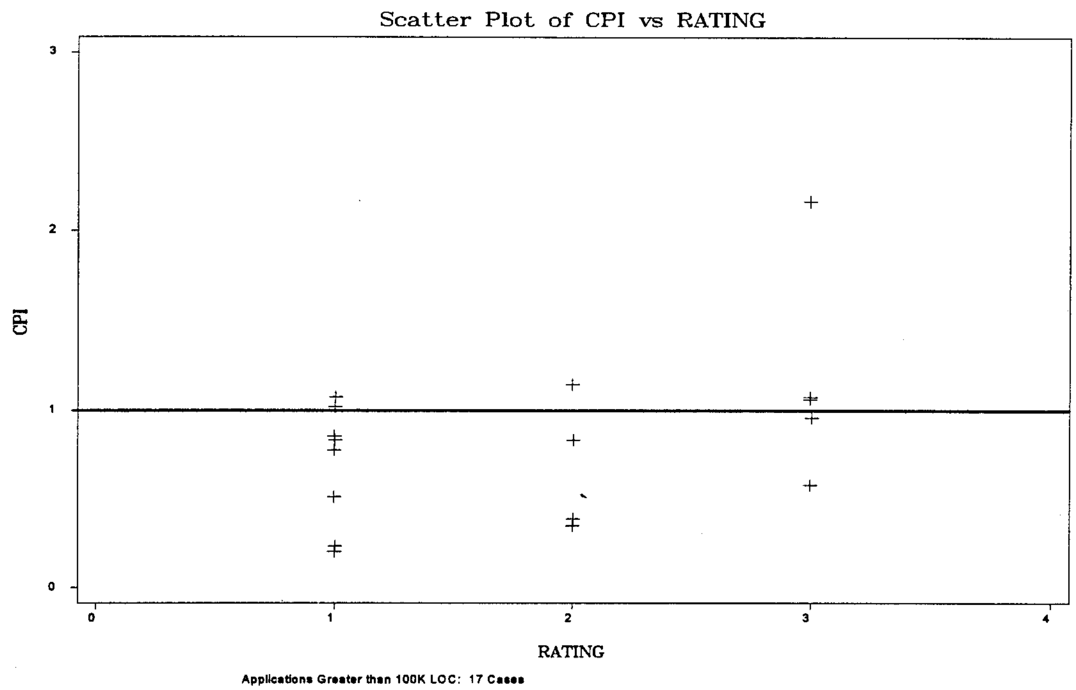

Figure K-3 Scatter Plot of CPI versus Rating for Applications Greater than 100K LOC

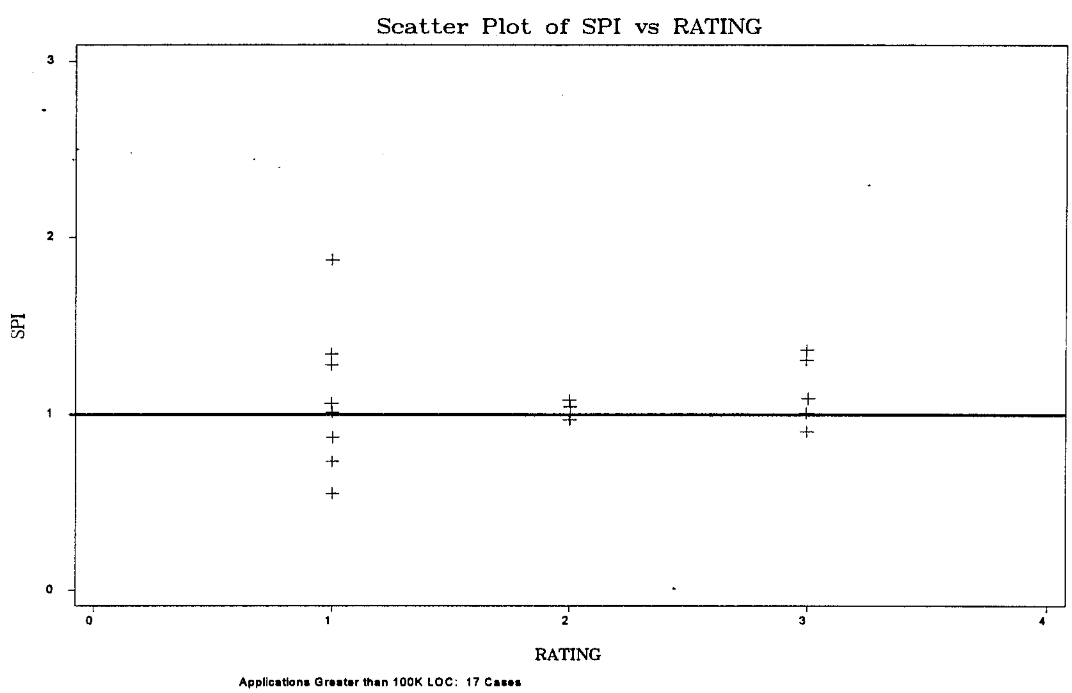

Figure K-4 Scatter Plot of SPI versus Rating for Applications Greater than 100K LOC

2. Box and Whisker Plots of CPI and SPI

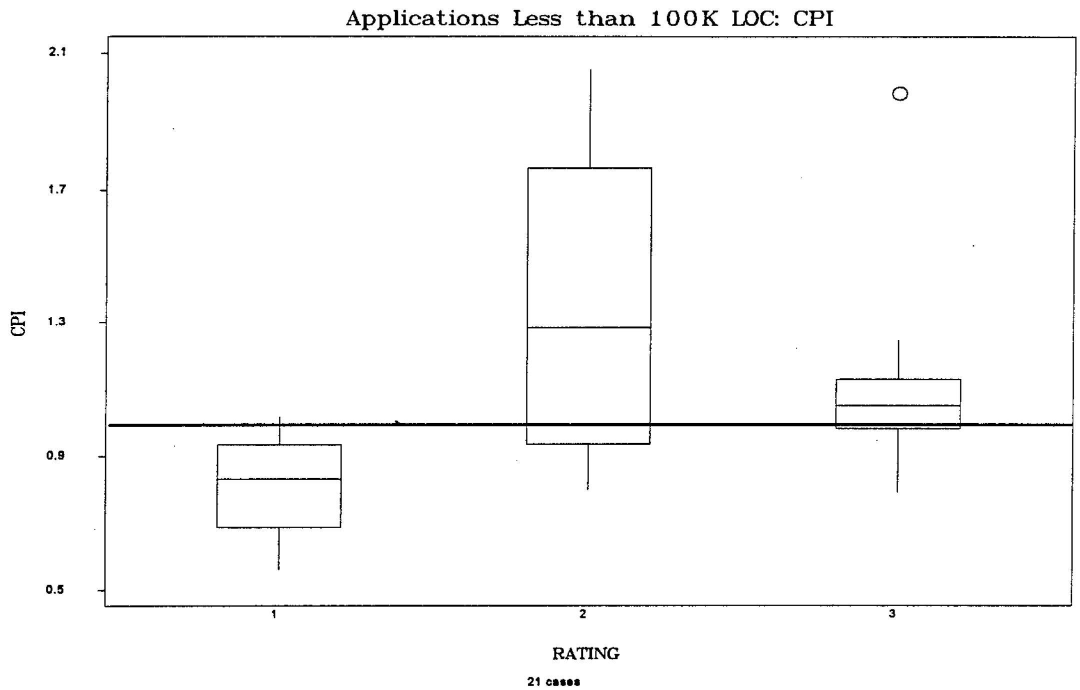

Figure K-5 Box and Whiskers Plot for CPI for Applications Less than 100K (LOC)

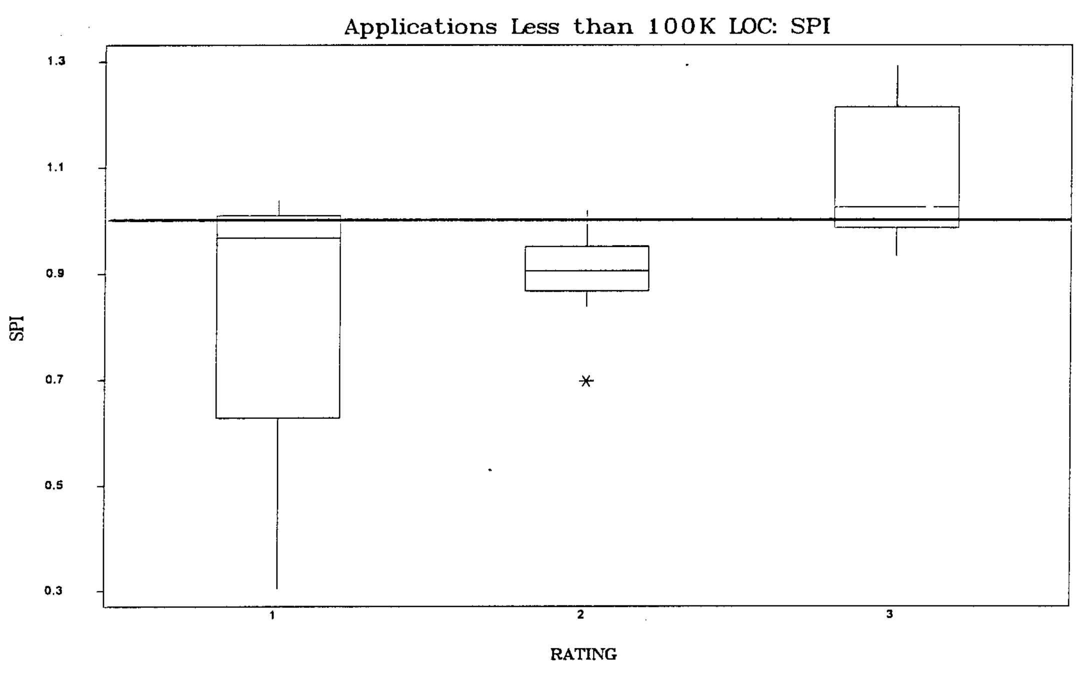

Figure K-6 Box and Whiskers Plot for SPI for Applications Less than 100K (LOC)

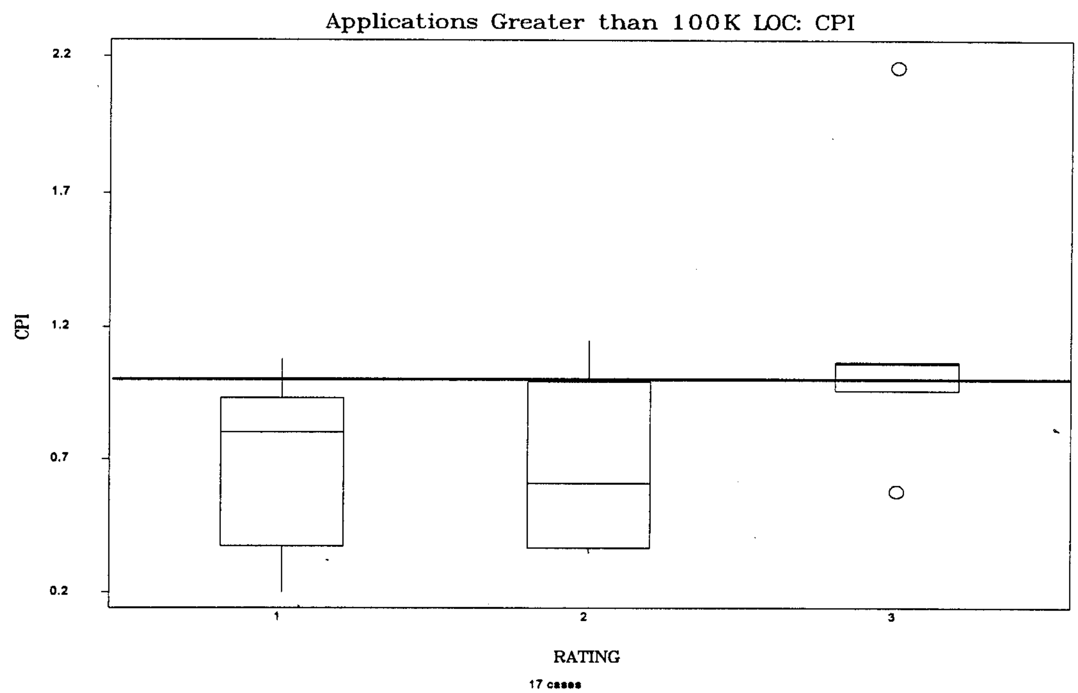

Figure K-7 Box and Whiskers Plot for CPI for Applications Greater than 100K LOC

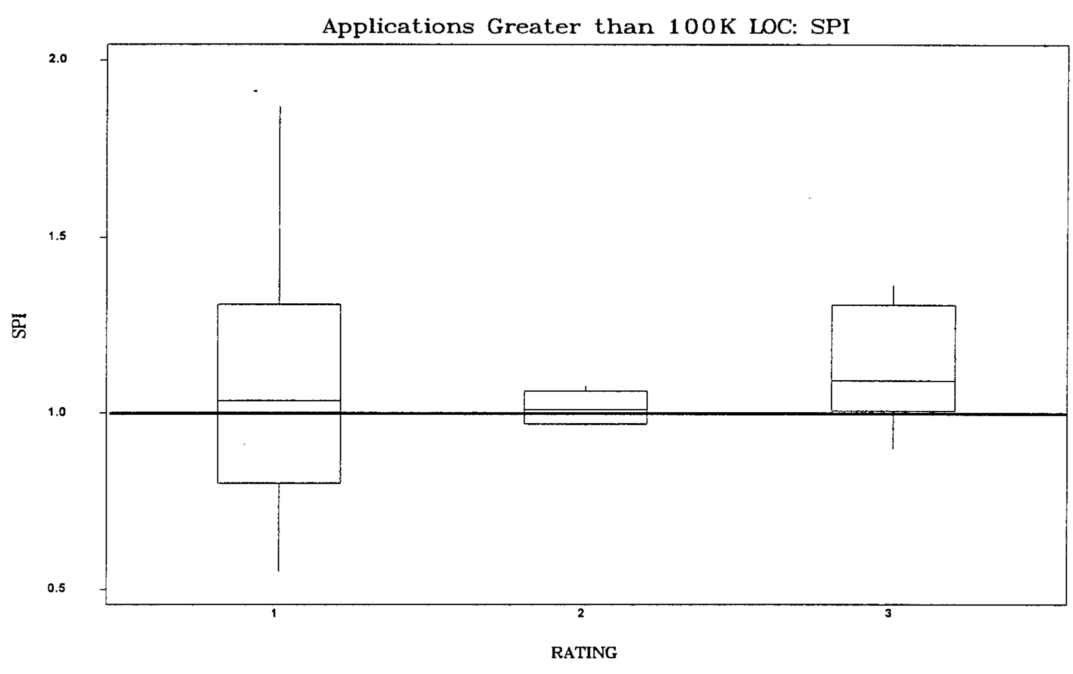

Figure K-8 Box and Whiskers Plot for SPI for Applications Greater than 100K LOC

3. Kruskal-Wallis and Multiple Comparison Tests

Table K-1
Kruskal-Wallis for CPI for Applications Less than 100K LOC

```
KRUSKAL-WALLIS ONE-WAY NONPARAMETRIC AOV FOR CPI BY RATING

            MEAN   SAMPLE
 RATING     RANK   SIZE
      1      5.5      4
      2     13.5      8
      3     11.2      9
TOTAL       11.0     21

KRUSKAL-WALLIS STATISTIC              4.4531
P-VALUE, USING CHI-SQUARED APPROXIMATION     0.1079
```

Table K-2
Multiple Comparison Matrix for CPI for Applications Less than 100K LOC

			Rating		
Rating	n	Mean Rank	1	2	3
1	4	5.5		--	--
2	8	13.5	1.031		--
3	9	11.2	-1.138	-3.23	

K-W Statistic of 4.4531, P=0.1079

Table K-3
Kruskal-Wallis for SPI for Applications Less than 100K LOC

```
KRUSKAL-WALLIS ONE-WAY NONPARAMETRIC AOV FOR SPI BY RATING

            MEAN   SAMPLE
 RATING     RANK   SIZE
      1      9.5      4
      2      6.7      8
      3     15.4      9
TOTAL       11.0     21

KRUSKAL-WALLIS STATISTIC              8.6046
P-VALUE, USING CHI-SQUARED APPROXIMATION     0.0135
```

Table K-4
Multiple Comparison Matrix for SPI for Applications Less than 100K LOC

			Rating		
Rating	n	Mean Rank	1	2	3
1	4	9.5		--	--
2	8	6.7	-4.169		--
3	9	15.4	3.17	-0.938	

K-W Statistic of 8.6046, P=0.00135

Table K-5
Kruskal-Wallis for CPI for Applications Greater than 100K LOC

```
KRUSKAL-WALLIS ONE-WAY NONPARAMETRIC AOV FOR CPI BY RATING

          MEAN   SAMPLE
 RATING   RANK   SIZE
      1    7.5      8
      2    8.0      4
      3   12.2      5
TOTAL      9.0     17

KRUSKAL-WALLIS STATISTIC              2.8706
P-VALUE, USING CHI-SQUARED APPROXIMATION    0.2380
```

Table K-6
Multiple Comparison Matrix for CPI for Applications Greater than 100K LOC

			Rating		
Rating	n	Mean Rank	1	2	3
1	8	7.5		--	--
2	4	8.0	-5.171		--
3	5	12.2	-.058	-2.013	

K-W Statistic of 2.8706, P=0.2380

Table K-7
Kruskal-Wallis for SPI for Applications Greater than 100K LOC

```
KRUSKAL-WALLIS ONE-WAY NONPARAMETRIC AOV FOR SPI BY RATING

          MEAN   SAMPLE
 RATING   RANK   SIZE
      1    8.6      8
      2    7.8      4
      3   10.6      5
TOTAL      9.0     17

KRUSKAL-WALLIS STATISTIC              0.7912
P-VALUE, USING CHI-SQUARED APPROXIMATION    0.6733
```

Table K-8
Multiple Comparison Matrix for SPI for Applications Greater than 100K LOC

			Rating		
Rating	n	Mean Rank	1	2	3
1	8	8.6		--	--
2	4	7.8	-4.871		--
3	5	10.6	-3.28	-3.413	

K-W Statistic of 0.7192, P=0.6733

4. *Descriptive Statistics*

Table K-9
Descriptive Statistics for Applications Less than 100K LOC

	Rating=1 CPI	Rating=1 SPI	Rating=2 CPI	Rating=2 SPI	Rating=3 CPI	Rating=3 SPI
N	4	4	8	8	9	9
Mean	0.8113	0.8193	1.3524	0.8959	1.1245	1.0916
Std Dev	0.1884	0.3462	0.4702	0.0978	0.3538	0.1333
Min	0.5623	0.3028	0.8000	0.6978	0.7924	0.9335
Median	0.8322	0.9681	1.2830	0.9060	1.0498	1.0262
Max	1.0187	1.0384	2.0506	1.0199	1.9819	1.2920
MAD	0.1018	0.0424	0.4141	0.0388	0.0774	0.0926
Skew	-0.3726	-1.1186	0.2283	-0.8603	1.6647	0.2402

Table K-10
Descriptive Statistics for Applications Greater than 100K LOC

	Rating=1 CPI	Rating=1 SPI	Rating=2 CPI	Rating=2 SPI	Rating=3 CPI	Rating=3 SPI
N	8	8	4	4	5	5
Mean	0.6875	1.0887	0.6801	1.0159	1.1659	1.1343
Std Dev	0.3347	0.4102	0.3825	0.0551	0.5903	0.1976
Min	0.2019	0.5507	0.3496	0.9660	0.5808	0.8998
Median	0.8023	1.0347	0.6115	1.0101	1.0611	1.0919
Max	1.0788	1.8676	1.1479	1.0774	2.1602	1.3652
MAD	0.2443	0.2727	0.2436	0.0407	0.1011	0.1921
Skew	-0.4432	0.6272	0.3073	0.1372	1.0485	0.0896

5. Multiple Comparison Calculations for Applications Less than 100K LOC

The Kruskal-Wallis H Test Statistic

$$H := \frac{12}{N\cdot(N+1)}\cdot\sum_{j=1}^{k}\frac{\left[R_j - \frac{n_j\cdot(N+1)}{2}\right]^2}{n_j}$$

Multiple Comparison Inequality

$$\left|\overline{R_i} - \overline{R_j}\right| \le z\cdot\sqrt{\frac{N\cdot(N+1)}{12}\cdot\left(\frac{1}{n_i}+\frac{1}{n_j}\right)}$$

where
k = number of samples
R_i, R_j = sum of the ranks in the ith and jth sample, respectively
n_i, n_j = number of observations in the ith and jth sample, respectively
N = total number of observations

Mean Rank of Ratings 1, 2, & 3 for CPI
$R_{cpi} := (5.5\ \ 13.5\ \ 11.2)$

Mean Rank of Ratings 1, 2, & 3 for SPI
$R_{spi} := (9.5\ \ 6.7\ \ 15.4)$

Number of Observations in Ratings 1, 2, & 3
$n := (4\ \ 8\ \ 9)$

Calculation of N $\quad N := n_{1,1} + n_{1,2} + n_{1,3} \quad N = 21$

Critical z Value - Corresponds to a .2 level of significance and provides at least a 80% level of confidence.
$z := 1.834$

<u>Multiple Comparisons</u> - "which populations differ from which others"
- If the difference in mean rank (DMR) between 2 samples is greater than the right-hand side (RHS) of the multiple comparison inequality above, there is a statistically significant difference between the 2 samples.
- The direction of the significant difference is determined by noting for each pair which sample has the larger sample mean.

Comparision of Rating 1 and 2

$DMR_{cpi} := \left|R_{cpi_{1,1}} - R_{cpi_{1,2}}\right| \quad DMR_{cpi} = 8$

$DMR_{spi} := \left|R_{spi_{1,1}} - R_{spi_{1,2}}\right| \quad DMR_{spi} = 2.8$

$RHS := z\cdot\sqrt{\frac{N\cdot(N+1)}{12}\cdot\left(\frac{1}{n_{1,1}}+\frac{1}{n_{1,2}}\right)} \quad RHS = 6.969$

For CPI $DMR_{cpi} - RHS = 1.031$ — CPI at Rating 2 > CPI at Rating 1 <==
For SPI $DMR_{spi} - RHS = -4.169$ — no significant difference <==

Comparision of Rating 2 and 3

$DMR_{cpi} := \left|R_{cpi_{1,2}} - R_{cpi_{1,3}}\right| \quad DMR_{cpi} = 2.3$

$DMR_{spi} := \left|R_{spi_{1,2}} - R_{spi_{1,3}}\right| \quad DMR_{spi} = 8.7$

$RHS := z\cdot\sqrt{\frac{N\cdot(N+1)}{12}\cdot\left(\frac{1}{n_{1,2}}+\frac{1}{n_{1,3}}\right)} \quad RHS = 5.53$

For CPI $DMR_{cpi} - RHS = -3.23$ — no significant difference <==
For SPI $DMR_{spi} - RHS = 3.17$ — SPI at Rating 3 > SPI at Rating 2 <==

Comparision of Rating 1 and 3

$DMR_{cpi} := \left|R_{cpi_{1,1}} - R_{cpi_{1,3}}\right| \quad DMR_{cpi} = 5.7$

$DMR_{spi} := \left|R_{spi_{1,1}} - R_{spi_{1,3}}\right| \quad DMR_{spi} = 5.9$

$RHS := z\cdot\sqrt{\frac{N\cdot(N+1)}{12}\cdot\left(\frac{1}{n_{1,1}}+\frac{1}{n_{1,3}}\right)} \quad RHS = 6.838$

For CPI $DMR_{cpi} - RHS = -1.138$ — no significant difference <==
For SPI $DMR_{spi} - RHS = -0.938$ — no significant difference <==

Figure K-9 Calculations for Multiple Comparison Test for Applications Less than 100K LOC

6. Multiple Comparison Calculations for Applications Greater than 100K LOC

The Kruskal-Wallis H Test Statistic

$$H := \frac{12}{N \cdot (N+1)} \cdot \sum_{j=1}^{k} \frac{\left[R_j - \frac{n_j \cdot (N+1)}{2}\right]^2}{n_j}$$

Multiple Comparison Inequality

$$\left|\overline{R_i} - \overline{R_j}\right| \le z \cdot \sqrt{\frac{N \cdot (N+1)}{12} \cdot \left(\frac{1}{n_i} + \frac{1}{n_j}\right)}$$

where
k = number of samples
R_i, R_j = sum of the ranks in the ith and jth sample, respectively
n_i, n_j = number of observations in the ith and jth sample, respectively
N = total number of observations

Mean Rank of Ratings 1, 2, & 3 for CPI
$R_{cpi} := (7.5\ \ 8.0\ \ 12.2)$

Mean Rank of Ratings 1, 2, & 3 for SPI
$R_{spi} := (8.6\ \ 7.8\ \ 10.6)$

Number of Observations in Ratings 1, 2, & 3
$n := (8\ \ 4\ \ 5)$

Calculation of N $\quad N := n_{1,1} + n_{1,2} + n_{1,3} \quad N = 17$

Critical z Value - Corresponds to a .2 level of significance and provides at least a 80% level of confidence.
$z := 1.834$

Multiple Comparisons - "which populations differ from which others"
- If the difference in mean rank (DMR) between 2 samples is greater than the right-hand side (RHS) of the multiple comparison inequality above, there is a statistically significant difference between the 2 samples.
- The direction of the significant difference is determined by noting for each pair which sample has the larger sample mean.

Comparision of Rating 1 and 2

$DMR_{cpi} := \left|R_{cpi_{1,1}} - R_{cpi_{1,2}}\right| \quad DMR_{cpi} = 0.5$

$DMR_{spi} := \left|R_{spi_{1,1}} - R_{spi_{1,2}}\right| \quad DMR_{spi} = 0.8$

$RHS := z \cdot \sqrt{\frac{N \cdot (N+1)}{12} \cdot \left(\frac{1}{n_{1,1}} + \frac{1}{n_{1,2}}\right)} \quad RHS = 5.671$

For CPI $DMR_{cpi} - RHS = -5.171$ — no significant difference <==
For SPI $DMR_{spi} - RHS = -4.871$ — no significant difference <==

Comparision of Rating 2 and 3

$DMR_{cpi} := \left|R_{cpi_{1,2}} - R_{cpi_{1,3}}\right| \quad DMR_{cpi} = 4.2$

$DMR_{spi} := \left|R_{spi_{1,2}} - R_{spi_{1,3}}\right| \quad DMR_{spi} = 2.8$

$RHS := z \cdot \sqrt{\frac{N \cdot (N+1)}{12} \cdot \left(\frac{1}{n_{1,2}} + \frac{1}{n_{1,3}}\right)} \quad RHS = 6.213$

For CPI $DMR_{cpi} - RHS = -2.013$ — no significant difference <==
For SPI $DMR_{spi} - RHS = -3.413$ — no significant difference <==

Comparision of Rating 1 and 3

$DMR_{cpi} := \left|R_{cpi_{1,1}} - R_{cpi_{1,3}}\right| \quad DMR_{cpi} = 4.7$

$DMR_{spi} := \left|R_{spi_{1,1}} - R_{spi_{1,3}}\right| \quad DMR_{spi} = 2$

$RHS := z \cdot \sqrt{\frac{N \cdot (N+1)}{12} \cdot \left(\frac{1}{n_{1,1}} + \frac{1}{n_{1,3}}\right)} \quad RHS = 5.28$

For CPI $DMR_{cpi} - RHS = -0.58$ — no significant difference <==
For SPI $DMR_{spi} - RHS = -3.28$ — no significant difference <==

Figure K-10 Calculations for Multiple Comparison Test for Applications Greater than 100K LOC

Bibliography

Analytical Software. Users Manual, Statistix Version 4.0. St Paul MN, 1992.

Basili, Victor R. and John D. Musa, "The Future Engineering of Software: A Management Perspective," IEEE Computer, 90-96 (September 1991).

Bersoff, Edward H. and Alan M. Davis. "Impacts of Life Cycle Models on Software," Communications of the ACM, 34: 104-118 (August 1991).

Besselman, Joseph J., Paul Byrnes, Cathy J. Lin, Mark C. Paulk and Rajesh Puranik. "Software Capability Evaluations: Experiences from the Field," SEI Technical Review '93: 1-30 (1993).

Boehm, Barry, "A Spiral Model of Software Development and Enhancement," IEEE Computer, 61-72 (May 1988).

Bollinger, Terry B. and Clement McGowan. "A Critical Look at Software Capability Evaluations," IEEE Software, 8: 25-41 (July 1991).

Carleton, Anita D., Robert E. Park, and Wolfhart B. Goethert. "Measurement Definitions for DoD Systems: Recommendations for an Initial Core Set," SEI Technical Review '93: 1-35 (1993).

Christensen, David S., and Scott R. Heise. "Cost Performance Index Stability," National Contract Management Journal, 25: 7-15 (1993).

Christensen, David S. Professor, Air Force Institute of Technology, Wright-Patterson AFB OH. Personal interview. 9 November 1994.

Davis, Alan M. "Operational Prototyping: A New Development Approach," IEEE Software, 70-78 (September 1992).

Department of the Air Force. Software Development Capability Evaluation, Vol 1. AFMCP 800-61. Wright-Patterson AFB OH: HQ/AFMC, 24 November 1993.

Department of the Air Force. Software Development Capability/Capacity Review. ASCP 800-5. Wright-Patterson AFB OH: HQ/ASC, 11 September 1992.

Department of Defense. Defense System Software Development. DOD-STD-2167. Washington: SPAWAR, 29 February 1988.

Devore, Jay L. Probability and Statistics for Engineering and Science (Third Edition). Pacific Grove CA: Brooks/Cole Publishing Company, 1982.

Dion, Raymond. "Process Improvement and the Corporate Balance Sheet," IEEE Software, 10: 28-35 (July1993).

Gibbons, Jean Dickinson. Nonparametric Methods for Quantitative Analysis. Chicago: Holt Rinehart and Winston. 1976.

Hersh, Art. "Where's the Return of Process Improvement?," IEEE Software, 10: 12 (July1993).

Honour Werth, Laurie. "Lecture Notes on Software Process Improvement," CMU/SEI-93-EM-8, (February 1993), AD-A265200.

Humphrey, Watts S., David H. Kitson, and Tim C. Kasse. 'The State of Software Engineering Practice: A Preliminary Report," CMU/SEI-89-TR-1, (February 1989), AD-A206573.

Humphrey, Watts S. Managing the Software Process. Reading MA: Addison-Wesley, 1989.

Humphrey, Watts S., Terry R. Snyder, and Ronald R. Willis, "Software Process Improvement at Hughes Aircraft," IEEE Software, 8: 11-23 (July 1991).

Lai, Robert, "The Move to Mature Processes," IEEE Software, 10: 14-17, (July1993).

Mosemann II, Lloyd K. "Improving Software Quality Through Measurement," CrossTalk, The Journal of Defense Software Engineering, 5: 2-5 (September 1992).

-----. "Why the New Metrics Policy," CrossTalk, The Journal of Defense Software Engineering, 7: 3 (April 1994).

Nicholas, John, M. Managing Business and Engineering Projects. Englewood Cliffs NJ: Prentice Hall, 1990.

Paulk, Mark C., Bill Curtis, Mary Beth Chrissis, and Charles V. Weber. "Capability Maturity Model, Version 1.1," IEEE Software, 10: 19-27 (July 1993).

Weinberg, Gerald M., Quality Software Management: Volume 1 Systems Thinking. New York: Dorset House Publishing, 1992.

Vita - Captain Robert M. Flowe

Captain Flowe was born in Alexandria, Virginia, on 1 April, 1960. He attended W.T. Woodson High School in Fairfax, Virginia, graduating in 1978. After high school, Captain Flowe attended Virginia Polytechnic Institute and State University in Blacksburg, Virginia, graduating with a Bachelor of Science in Aerospace and Ocean Engineering in June, 1984. While pursuing his undergraduate degree, Captain Flowe worked for the National Aeronautics and Space Administration at the Johnson Space Center, in Houston, Texas as a cooperative education engineering trainee.

While still an undergraduate, Captain Flowe enlisted in the Air Force on 13 September 1983. Upon graduation from Virginia Tech, he attended Officer Training School at Lackland Air Force Base, Texas, where he received his commission on 7 November, 1984. Upon commissioning, Captain Flowe was assigned to the Titan III Systems Program Office at Los Angeles Air Force Base, Los Angeles, California. During his five-year assignment to Los Angeles Air Force Base, Captain Flowe played a key role in the operation and management of the Titan 34D space launch system. Captain Flowe was reassigned in September 1989 to the 6595th Aerospace Test Group, Vandenberg Air Force Base, California, where Captain Flowe supported integration, test, and launch of Titan II and Titan IV space launch vehicles. Captain Flowe was selected to attend the Air Force Institute of Technology in May, 1993.

Permanent Address:
Robert M. Flowe
4305 Olley Lane
Fairfax, VA 22032

Vita - Captain James Thordahl

Captain Thordahl was born in Ayre, Scotland, on 4 December 1965. He attended Paraclete High School in Lancaster, California. He earned a Bachelor of Science degree in Aerospace Engineering from the University of Notre Dame, Notre Dame, Indiana, in May 1988. Upon graduation he was commissioned through Air Force ROTC and entered active duty in January 1989 as a Laser Weapons Project Officer assigned to the Phillips Laboratory, Kirtland AFB, New Mexico. While assigned as a Laser Weapons Project Officer, Captain Thordahl managed several programs in support of high energy laser development for the Ground Based Laser Antisatellite program, including the first ever coupled multiple-output Chemical Oxygen-Iodine Laser (COIL). In April 1991, he was selected to serve as the Air Force Maui Optical Site (AMOS) Program Manager where he was responsible for providing operational electro-optical data to the Space Surveillance Network and optical research support to all DoD services, National Laboratories, and the SDIO. In May of 1993, he entered the School of Systems and Logistics, Air Force Institute of Technology.

Permanent Address:
James B. Thordahl
42910 Chicory Avenue
Lancaster, CA 93534

REPORT DOCUMENTATION PAGE

Form Approved
OMB No 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	December 1994	Master's Thesis

4. TITLE AND SUBTITLE

A CORRELATIONAL STUDY OF THE SEI'S CAPABILITY MATURITY MODEL AND SOFTWARE DEVELOPMENT PERFORMANCE IN DOD CONTRACTS

5. FUNDING NUMBERS

6. AUTHOR(S)

Robert M. Flowe, Captain USAF
James B. Thordahl, Captain USAF

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Air Force Institute of Technology,
WPAFB OH 45433-6583

8. PERFORMING ORGANIZATION REPORT NUMBER

AFIT/GSS/LAR/94D-2

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

HQ AFMC/ENPC
Wright-Patterson AFB, OH 45433

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

The Software Engineering Institute's (SEI's) Capability Maturity Model (CMM) is widely used to measure an organization's software development process maturity. The Department of Defense (DoD) has adopted this model with the belief that a more mature software development process will result in a more successful software project. Although there is a growing body of anecdotal evidence supporting this presumed correlation, there is currently no empirical evidence. Thus, the goal of our research was to determine the nature of the correlation, if any, between software process maturity and software project success, where process maturity is based on a CMM rating and success is based on the parameters of cost and schedule. To investigate this correlation we identified software unique projects, obtained CMM rating information on the contractor, collected cost and schedule data from a time frame representative of the rating, and interviewed project personnel to collect project context information. Using plots of cost and schedule performance versus rating level and nonparametric statistical techniques we found that, within our dataset, a correlation does exist between software development process maturity and project performance. The nature of this correlation appears to be improved cost and schedule performance with higher software process maturity.

14. SUBJECT TERMS

Software, Process, Maturity, Improvement, Metrics
Management, SEI, CMM, Ada, C/SCSC

15. NUMBER OF PAGES

300

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	UL