Metrics for Measuring the Value of Computer Integrated Manufacturing (CIM) Systems

by

William Wofford Crandall, Jr.

A.B., Woodrow Wilson School of Public and International Affairs
Princeton University (1987)

Submitted to the Department of Electrical Engineering and Computer Science and to the MIT Sloan School of Management in Partial Fulfillment of the Requirements for the Degrees of

Master of Science in Electrical Engineering and Computer Science
and
Master of Science in Management

at the Massachusetts Institute of Technology
May 1994

© Massachusetts Institute of Technology, 1994. All rights reserved.

Signature of Author ____________________________

Department of Electrical Engineering and Computer Science
MIT Sloan School of Management
May 6, 1994

Certified by

John V. Guttag, Professor of Computer Science and Engineering
Thesis Supervisor

Certified by

Michael A. Cusumano, Professor of Management Science
Thesis Supervisor

Accepted by ____________________________

F. R. Morgenthaler
Department of Electrical Engineering and Computer Science
Chairman, Departmental Committee on Graduate Students

Accepted by ____________________________

Jeffrey Barks
Associate Dean, Sloan Master's and Bachelor's Programs
Acknowledgments

The author gratefully acknowledges the support and resources made available to him through the Leaders for Manufacturing Program, a partnership between MIT and major U.S. manufacturing companies.

The author would particularly like to thank the people at the Avalon Corporation whose ideas and energies made this thesis and the internship upon which it is based possible, including Stew Caulfield, Dave Decker, Frank Faron, Chuck Feltner, Dirk Heitzman, Lew Lyons, John O'Brien, and Barry Sobel.

The author would also like to thank his wife, Karen Crandall, for her love and help during the development of this work.
Metrics for Measuring the Value of Computer Integrated Manufacturing (CIM) Systems

by

William Wofford Crandall, Jr.

Submitted to the Department of Electrical Engineering and Computer Science and to the MIT Sloan School of Management on May 6, 1994 in partial fulfillment of the requirements for the degrees of Master of Science in Electrical Engineering and Computer Science and Master of Science in Management

ABSTRACT

Being able to measure the value created by Computer Integrated Manufacturing (CIM) systems is important because much of the literature indicates that investments in CIM and other forms of information technology are poorly correlated with improvements in productivity. Unfortunately, measuring the value of CIM is difficult. This makes it difficult to know whether or not a proposed CIM system will create value, if an existing system has created value, or, for those CIM systems that require new software, how the software development process should best be guided to produce systems of value.

In an attempt to solve these problems, the thesis explores different approaches to creating metrics to measure the value of CIM systems. Specifically, the thesis analyzes a metric defined at one of the MIT Leaders for Manufacturing Program's sponsor companies for measuring the value of CIM in dollar terms and explains why this is a poor metric for many CIM projects. The thesis then details an alternative approach for defining value metrics specific to individual CIM projects. The methodology proposed here emphasizes the production of metrics that can be used during the software development process to balance traditional productivity, schedule, cost, and quality software metrics. The thesis describes a test of this methodology conducted at LFM sponsor company.
Table of Contents

Acknowledgments .......................................................... 2
Abstract ............................................................................. 3
Table of Contents .......................................................... 4
Introduction ......................................................................... 5
Section 1: Why Is This Topic Important? ......................... 9
   Understanding the Productivity Paradox ......................... 10
   Metrics for Measuring Value ....................................... 16
Section 2: Characteristics of Good Metrics and Successful Metrics Programs ........................................... 20
   Characteristics of Good Metrics .................................. 20
   Characteristics of Successful Metrics Program ............... 24
Section 3: Value Metrics at the Avalon Corporation ......... 30
   Avalon's Systems Efficiency Program (SEP) ................ 30
   SEP's Plant Floor Action Team (PFAT) ....................... 31
   PFAT's Systems Request Effectiveness (SRE) Metric .... 32
   Dollar Value, a Poor Metric for Many CIM Projects .... 34
   Cost Savings ............................................................... 35
   Types of CIM Projects ................................................ 42
   Continued Use of the SRE Metric ............................... 46
   The March of Folly ...................................................... 55
   Other Attempts to Measure Value in Dollar Terms ....... 56
   Value Metrics at the Avalon Corp. ............................. 60
Section 4: A Methodology for Creating Value Metrics ....... 61
   Results Metrics .......................................................... 61
   Leading Metrics and the Software Development Process .. 62
   Process-Independent Methodology ............................. 63
   Overview of and Motivation for the Methodology .......... 65
   Use of the Methodology .............................................. 67
   Definition of the Methodology .................................... 68
   Step 1: Talk to Customers ......................................... 69
   Step 2: Organize and Prioritize Customer Needs .......... 73
   How to Create a KJ Diagram ...................................... 75
   Step 3: Verify Results with Customers ....................... 77
   Customer Value Survey: Sample Format ..................... 78
   Step 4: Think of Metrics ............................................ 79
   Step 5: Evaluate the Metrics ...................................... 82
   Questions to Be Answered in the Evaluation Matrix ....... 83
   Step 6: Select Key Metrics ........................................ 84
   Step 7: Verify Choices with Customers ....................... 87
   Step 8: Implement and Use ........................................ 88
   Step 9: Improve this Process ..................................... 89
   Advantages of the Methodology .................................. 90
   Disadvantages of the Methodology ......................... 91
   One CIM Project Team's Experience with the Methodology ......................................................... 92
   Details of the Test .................................................... 95
   Critique of the Test and the Methodology ................. 114
   Overall Comments on the Methodology .................... 125
Conclusion ........................................................................ 127
Bibliography ............................................................... 129
Introduction

"How do I know I'm doing the right thing?"
-- Chuck Feltner, Director, Manufacturing Planning & Information Systems, Components Operations, Avalon Corporation
April 24, 1993

Measuring the value of Computer Integrated Manufacturing (CIM) systems is difficult. CIM systems are the software and hardware systems used to control, coordinate, and track activity on a plant floor. While CIM systems can be integrated with business and engineering systems (such as payroll, order entry, cost accounting, and computer-aided design systems), the primary purpose of CIM systems is to either:

- direct the actions of machines and people in the plant, or
- provide the people in the plant with the information they need to direct the actions of machines and people and to analyze and improve the operations of the plant.

Examples of CIM systems include cell and area controllers, statistical process control (SPC) systems, automated materials handling systems, and assembly sequencing systems.

Measuring the value of CIM systems is difficult because managers are unable to readily quantify accurately the value of many of the most important benefits of these systems (such as improved quality, flexibility, and responsiveness). This forces managers to make important decisions about CIM systems using incomplete or flawed quantitative analyses, intuition, anecdotes, or politics. These are imperfect measures for guiding the behavior of CIM teams towards creating maximum value for the firm.

---

1 To protect information confidential to the firm at which the research for this thesis was conducted, the firm's name and the names of its operating divisions have been disguised.
During an internship at the Avalon Corporation sponsored by the MIT Leaders for Manufacturing Program, the author was asked to explore this issue and to find better metrics for measuring the value of CIM systems, with a particular emphasis on how these metrics could be used during the software development process to guide software development teams towards producing CIM systems that create maximum value. This thesis describes the results of this research and is broken into four sections:

1. Section 1 discusses why this topic is both important and a continuing subject of debate. First, the great bulk of the literature has found no link between investments in information technology and productivity. While recent research has contradicted this conclusion, there are still good reasons for managers to scrutinize more carefully the benefits that they expect to receive from CIM systems. Second, there are only a few value metrics in widespread use and they tend to have major flaws. The lack of good value metrics again should encourage managers to find better ways of measuring the value of CIM.

2. Section 2 summarizes the software engineering literature to define the characteristics of good metrics and successful metrics programs.

3. Section 3 analyzes Avalon's attempts to deploy metrics that measure the value of CIM systems in dollar terms. It describes the nature of the three types of savings generated by CIM systems and how these savings fall into the four types of CIM projects. It finds that for most CIM projects, the dollar amount of value is hard to calculate accurately in an easy-to-obtain manner. Therefore, metrics for measuring value in dollar terms fail to have the characteristics of good metrics and should not be used.

4. Section 4 proposes a methodology for developing project-specific value metrics, outlined in Figure 1. It describes the methodology and the thinking behind it in detail and then examines the experience of a CIM team at Avalon that used the methodology. The methodology emphasizes two things. First, it emphasizes the
development of results metrics that can assess the system's performance relative to value after installation and that can aid process improvement in the plant. Second, for CIM projects that require new software, it emphasizes the development of leading metrics that can be used to guide the CIM software development team towards creating value on the factory floor. Such value metrics can balance traditional software engineering metrics for productivity, schedule, cost, and quality; without value metrics in place, the development process may be unwittingly optimized for these other measures at the expense of value.

Overall, this discussion can aid managers in making better decisions about CIM systems by:

- suggesting ways of thinking more concretely about measuring the value of IT,
- helping them to understand why some value metrics, such as ones that try to measure value in dollar terms, frustrate so many people and why these metrics should be abandoned, and
- providing a methodology for creating their own project-specific value metrics that can be used as part of an effective metrics program to increase the value created by their CIM systems, especially by ensuring that the software development process pays more attention to the issue of value.

Managers should then have better tools to determine if they are "doing the right thing."
Figure 1: Methodology for Selecting Project-Specific Value Metrics

1. Talk to customers to find out what matters most to them in terms of value.
   TOOL TO USE
   Structured Interviews

2. Organize & prioritize those needs.
   KJ Diagram

3. Verify results with customers.
   Customer Survey

4. Think of metrics to measure those needs.
   Brainstorming

5. Evaluate metrics with a matrix.
   Evaluation Matrix

6. Select small group of key metrics.
   Selection Matrix

7. Verify choices with customers.
   Presentation

8. Implement and use.
   Metrics & Metrics Program

9. Improve this process based on experience.
   TQM Problem Solving Tools
Section 1: Why Is This Topic Important?

For a profit-maximizing firm, information technology (IT) creates value when it increase the present value of the firm's profits. By raising revenues, lowering expenses, or some combination thereof in excess of its cost, an IT system produces value for the firm. Given this definition, the study of the metrics for measuring the value of information technology (IT) in general and Computer Integrated Manufacturing (CIM) systems in particular is important for two reasons:

1. In spite of massive and growing investment in IT, the evidence is not persuasive that this investment has created value, a phenomenon known as the "productivity paradox."

2. Traditionally, the IT and software engineering communities have placed little attention on metrics for measuring value because these metrics are so difficult to create. Instead, numerous IT and software engineering metrics for cost, schedule, productivity, and quality (where quality is narrowly defined as the absence of defects) are described in the literature and used actively in industry. Unfortunately, it is possible to deliver CIM systems on-time, on-budget, and produced with high productivity and no defects, only to have them sit idle, unused, and unwanted on the factory floor, creating no value.

These facts should encourage managers to focus on improving their understanding of how IT can create value, what IT development processes are best suited for creating value, and how progress towards creating value can be measured. This thesis explores the last element of this list: creating metrics to measure the value of CIM systems as they are developed and deployed.

Note that the software needed for a new CIM system can come from many places: it can be developed from scratch in-house, contracted out to a third-party developer, bought off the shelf, customized by a systems integrator, and so on. The
value derived from the CIM system is independent of the software development process. But in the cases where fresh software is developed, it should be possible to correlate certain measurements of the development process with good outcomes in terms of value. As this thesis examines how the value of CIM can be measured, it focuses on how these measurements can be used in the software development process so that engineers will focus not only on developing defect-free software quickly and cheaply but on ensuring that value is created for their factory-floor customers. The absence of good metrics for establishing the value of a CIM system or guiding the progress of the development of such a system towards high value is one of the reasons that so many CIM systems fail to hold up to their promises of high value.

**Understanding the Productivity Paradox**

The term "productivity paradox" arose as researchers wrestled with the fact that while firms have been making ever-growing investments in IT, there has been no corresponding increase in total factor productivity to justify this investment.\(^2\) Fortunately for those making IT investments, recent work has persuasively argued that the productivity paradox has been resolved. But on the whole, the substantial body of evidence for this paradox, even if it no longer exists, should caution managers that investments in IT do not necessarily have a good guaranteed return and that they should spend more time understanding how and ensuring that their IT projects will create value.

The evidence of massive investment in IT is clear. Over the past two decades, firms have become much more capital intensive, especially manufacturing firms.

---

\(^2\) Productivity is the ratio of a firm's outputs to its inputs. Profit is the difference between outputs and inputs measured in dollar terms. In the business world, improvements in productivity generally mean improvements in profitability even though mathematically, this need not be true. Therefore, this thesis treats improvements in productivity and improvements in profitability interchangeably as definitions of "value."
Simultaneously, the share of new capital investment devoted to IT has increased sharply. [Thackenkary 1993] notes that from 1980 to 1989, U.S. expenditures for IT increased at a nearly 9% compound annual rate, far higher than for all other types of non-residential durable equipment, bringing the cumulative investment in IT to more than $1 trillion. In the fourteen American manufacturing industries surveyed by [Morrison and Berndt 1991], the average share of aggregate capital stock devoted to high-tech office and information technology equipment rose from 9.1% in 1976 to 32.1% in 1986. [Brynjolfsson and Hitt 1993] show that while annual spending on computer capital grew by about 85% in real terms between 1987 and 1991 in their 74 firm sample, non-computer capital spending grew only about 10% and employment dropped 2%. Given that the real cost of IT has fallen dramatically over the past twenty years as well, this increase in IT capital spending is all the more impressive.

The evidence of a return on this investment is less clear. Many studies have found no correlation between IT investment and productivity improvements while others have shown significant correlations only under certain circumstances. Specifically,

- [Morrison and Berndt 1991] found that the marginal benefit of IT investments in the fourteen American manufacturing industries they studied was less than the marginal cost, a return of only 80 cents on the dollar at the margin, indicating an overinvestment in IT in these industries as a whole.

- Later work by the same authors, [Berndt and Morrison 1992], found very limited evidence of a positive relationship between profitability and the ratio of high-tech capital to total capital. They also found evidence, confirmed in another study, [Berndt et al. 1992], that increases in the ratio of high-tech capital to total capital are negatively correlated with average labor productivity. That is, additional IT capital costs firms more labor than it saves. No conclusions on IT investment's effects on total factor productivity were drawn however.
• Using five different econometric models, [Loveman 1988] found no increase in output attributable to investment in IT for any of the models in the data set he studied. The data was drawn from about sixty manufacturing business units from about twenty large American and Western European manufacturing firms between 1978 and 1984. Loveman's attempts to find significant correlations between IT investment and output by segmenting the data (specifically, by separating low-IT-intensity firms from high ones, durables manufacturers from non-durables ones, and market share leaders from laggards) were also unsuccessful.

• [Weill 1990] and [Weill 1992] studied 33 manufacturing firms in the American valve industry from 1982 to 1987 but found "no significant relationship between total IT investment in previous years and any measure of 1987 firm performance" [Weill 1990, p. 14]. Yet by breaking down IT investment into three categories, he found that heavy investments in transactional IT (i.e., systems for processing the transactions of the firm, a type of investment often intended to reduce direct labor) did correlate significantly with return on assets and labor productivity. No such relationship was found for the other type types of IT: informational (i.e., systems for supporting management control, planning, communication, accounting, etc.) and strategic (i.e., systems for gaining competitive advantage and increasing market share, such as sales support systems).

• Further, in a review of the literature, [Weill 1990] cites two studies of the banking industry that found little relationship between performance and IT expenditures. He cites three other studies that together indicate there may be an optimum level of IT investment depending on a firm's profitability and position within an industry.

• Roach's analysis of the American service sector found that "massive investments in technology [by service firms] simply have not improved productivity" [Roach 1988, p. 85]. His data show a huge, sustained increase in the amount of IT capital per white-collar worker in the period 1962 to 1989 while white-collar productivity was
up little over the same time. This is especially worrisome given that, according to Roach, the American service sector is investing more than $100 billion in IT annually and now owns more than 85% of installed IT capital. Unfortunately, since Roach was writing for a wide audience, the source of his data and his methodology are not as clear as other more academic studies, making it difficult to critique his analysis.

- A study by [Siegel and Griliches 1991] found a positive and significant correlation between productivity growth and IT investment but the authors did not state "whether this result reflects errors of measurement of capital or is, in fact, indicative of the importance of computers as a determinant of productivity growth" [Siegel and Griliches 1991, p. 24]. In fact, most of this paper is devoted to understanding how errors in measurement of some manufacturing inputs (specifically, purchased services, outsourcing, and IT) occur and what effect they have on measuring manufacturing output. With respect to IT, they note the sustained, large drops in real computer prices from 1969 to 1987 and then state that "[g]iven the large increase in nominal expenditures on computers during this period, real investment in these machines and their relative weight in the capital stocks of representative industries are also substantially higher. Important technological improvements embodied in successive generations of computers may not have been properly accounted for in inventory deflators for these capital goods" [Siegel and Griliches 1991, p. 3]. The difficulty of making accurate, meaningful measurements of the size of IT capital investments is a theme repeated frequently in all of these studies and is explored further below.

Together, these studies define the productivity paradox: massive investment, little apparent return.

The existence of such a paradox troubles researchers. As Bandt and Morrison note:
For economists studying rational choices made by profit-maximizing firms, it is somewhat difficult at best to conclude that, like lemmings gone to sea, U.S. manufacturing firms have irrationally over-invested in IT capital. [Berndt and Morrison 1992, p. 19]

Fortunately for those making investments in IT, new work has raised doubts about the productivity paradox. For example, recent firm-level work by [Lichtenberg 1993] found that there are substantial excess returns to both IT capital and labor investment. His analysis also indicates that the average marginal rate of substitution between IT and non-IT employees is six, meaning that "one [IT] employee can be substituted for six non-[IT] employees without affecting output" (although the wages of IT employees tend to be much higher than those of non-IT employees) [Lichtenberg 1993, p. i].

Unfortunately, Lichtenberg does not examine in detail why his research contradicts the results cited above.

Perhaps the best research to date contradicting the productivity paradox was conducted by [Brynjolfsson and Hitt 1993]. They make a strong case that the productivity paradox has been resolved and has shed insight into why the notion of a productivity paradox arose in the first place. In their study of 380 American firms in 48 industries, representing $2.1 trillion in 1991 revenues or about half of U.S. GDP, they found that:

information systems have made a substantial and statistically significant contribution to firm output. We find that between 1987 and 1991, return on investment (ROI) for computer capital averaged 54% in manufacturing and 68% for manufacturing and services combined in our sample. [Brynjolfsson and Hitt 1993, p. i]

Because their methodology was very similar to that used by Loveman and by Berndt and Morrison, they developed several explanations for why their results were different. These explanations were codified in [Brynjolfsson 1993]. There, Brynolfsson explores the history of the productivity paradox and gives four answers to the question "Why haven't computers measurably improved productivity?":

14
1. **Measurement error:** Outputs (and inputs) of information-using industries are not being properly measured by conventional approaches.

2. **Lags:** Time lags in the payoffs to IT make analysis of current costs versus current benefits misleading.

3. **Redistribution:** It is especially likely that IT is used in redistributive activities among firms, making it privately beneficial without adding to total output.

4. **Mismanagement:** The lack of explicit measures of the value of information makes it particularly vulnerable to misallocation and overconsumption by managers. [Brynjolfsson 1993, p. 76]

In light of these answers, Brynjolfsson and Hitt give the following explanations for the differences in their results:

First, with regard to measurement error, their analysis techniques and the size of their data set helped overcome many of the problems with errors in the data that are acknowledged in the other studies.

Second, with regard to lags, Brynjolfsson and Hitt examined a later time period than the other studies. Because the firms' massive investments in IT were relatively recent, the absolute amount of IT capital may have been too small earlier to make its effect on overall productivity noticeable. Also, the firms may have experienced learning curve effects. As they first deployed IT, the firms may have gotten immediate benefits smaller than their costs but they would have also gained learning that would help them increase the benefits derived from later investments. If, in the long run, the learning helped produce total benefits that exceeded total costs, appropriately discounted, this would have been a rational strategy for a profit-maximizing firm.

Third, with regard to redistribution, they argue that they were able to use firm-level data instead of industry-level data and that the breadth and quality of their firm-level data was better than that used by anyone else before. They believe firm-level data can be more revealing than industry-level data because:

> if consumers value a benefit like variety, then they are likely to shift their purchases away from firms offering only mass-produced products toward firms offering more customized products. Because the cost per
unit at the customizing firm may be higher, it may appear to have lower productivity by conventional measures if similar products are compared. However, the increased sales by the customizing firm would indicate that they are adding real value to their products. Note that while the benefits of spending (and the costs of not spending) would show up as increases in sales [at] the firm level, industry revenues as a whole do not necessarily increase, so data on aggregate industry revenues could be misleading. [Brynjolfsson and Hitt 1993, p. 3]

This thinking echoes the work of [Brooke 1992], who makes the argument that IT "has altered the economics of production in favor of differentiated output, and that our methods of productivity measurement tend to discount the benefits of greater product variety" [Brooke 1992, p. i]. Brooke examined six sectors of the U.S. economy from 1950 to 1989 and found a sharp rise in product differentiation and IT adoption, especially after 1973, but only slow growth in productivity. He attributes the productivity paradox to the measurement problems inherent to such a sector-level view. Brynjolfsson and Hitt's firm-level view avoids this problem.

While all of Brynjolfsson and Hitt's arguments help explain why they did not find evidence of a continuing productivity paradox, Brynjolfsson singles out measurement error as the most significant cause of the productivity paradox. Nonetheless, the overall body of evidence and the lags, redistribution, and mismanagement answers provided by Brynolfsson should caution managers that a healthy return is not easily guaranteed for investments in information technology. The data indicate that managers would be well advised to place greater attention on specifying what value is to be derived from IT projects and how progress towards that value is to be measured.

**Metrics for Measuring Value**

Unfortunately, there has not been much success to date in developing good metrics for measuring the value of IT in general and of CIM systems in particular. The
two types of IT value metrics in common use, financial calculations and customer satisfaction metrics, both suffer from serious drawbacks.

Financial calculations (such as return on investment, payback period, and net present value) are common measures of the value of IT projects. However, as is well noted in the literature, it is often difficult to quantify in dollar terms many of the most important benefits of IT for easy inclusion in those calculations. For CIM systems, these benefits include improved flexibility, responsiveness, speed, and so on. Even in organizations where financial calculations are the basis for IT investment decisions, this difficulty encourages managers to justify IT investments "by faith alone," notes [Kaplan 1986], by "the use of heuristics, rather than strict cost/benefit accounting," states [Brynjolfsson 1993, p. 75], or by the use of anecdotes and political maneuvering. Such techniques inadequately and often inaccurately measure the value of IT and CIM projects and lead to suboptimal investments. All of this will be discussed in greater detail in Section 3.

The most common customer satisfaction metric is a customer satisfaction rating. These ratings tend to be simple rankings on a scale of one to five of overall satisfaction. [Marion 1993] cites a Coopers & Lybrand study of more than 200 chief executive officers and other senior managers which found that 93% use customer satisfaction as a primary measure of the performance of the chief information officer. A Gartner Group report notes that chief executive officers, chief financial officers, and chief operating officers place user satisfaction ahead of project financials, return on investment, competitive necessity, and gut feel as the best method for measuring IT's effectiveness [Rosser and Small 1992].

The motivation behind using customer satisfaction ratings is that, since quantifying the value of IT is tough, the only way to know if you are doing the right thing is to ask your customer. If the customer is happy, ipso facto, you are creating value.
But there are serious problems with using a customer satisfaction rating:

1. **Lagging**: A rating tells the IT organization how it has done but gives little indication of how it is going to do. In particular, for IT projects that include the development of new software, such a lagging metric gives no guidance on what should or should not be done during the development process to create more value or produce higher satisfaction.

2. **Infrequent**: Because customer satisfaction surveys are generally conducted infrequently, often only once a year, it can take a long time to notice problems arising or improvements taking effect. Infrequency can also lead to lack of visibility and disuse. Infrequency can also make the results inaccurate in that they may reflect customer satisfaction only over the last week or month, not over the last year.

3. **Aggregate**: Generally conducted at division level or higher, a rating only gives feedback on an IT organization's overall performance and reveals little about individual projects or systems. Also, because the average rating typically is relatively high ([Marion 1993] notes that an 80-85% positive customer satisfaction rating is the norm), managers are often content with the rating and do little to improve it or to identify the unsatisfied parts of the customer base and address their concerns.

4. **Misleading**: Some challenge the notion that if the customer is happy, IT is necessarily creating value. If customers are inwardly focused and are unaware of or are unable to calculate the costs and benefits of their IT systems, just because customers are satisfied does not mean that the benefit generated (in terms of higher revenues and lower expenses) exceeds the cost of the investment nor does it mean that the maximum value or return possible has been achieved.

Other types of customer satisfaction measures are less widespread. In a series of IT benchmarking studies against leading U.S. and Japanese firms in 1991 and 1992,
Avalon found that only five of the eleven firms studied had customer satisfaction metrics beyond a simple customer satisfaction rating [Avalon 1993d]. The metrics included defect repair time, online response time, ease of use, percentage of commitments met, etc. While all of these metrics -- as well as generalized productivity, cost, schedule, and quality/defect metrics -- can be components of satisfaction, they do not get directly at the question of value.

Overall, the most commonly-used financial calculations and customer satisfaction metrics are, at best, imperfect measures of the value of IT and CIM systems. This absence of good value metrics plus the presence of an alleged "productivity paradox" steers one away from concluding that IT and CIM investment necessarily create value. This fact drives this thesis to consider more deeply if there are other ways of measuring the value of IT and specifically CIM systems and then to try to tie those measurements back into the software development process. To do so, the characteristics of good software engineering metrics and successful software metrics programs must first be defined.
Section 2: Characteristics of Good Metrics and Successful Metrics Programs

Here we examine the characteristics of good software metrics and successful software metrics programs. The use of metrics is widespread in the software industry and much has been written on the subject. While there is little scientific analysis of what defines a good metric and a successful metrics program, there is a remarkable degree of consensus in the literature written both by academics and industry practitioners on their practical experiences as to what is needed to create a good metric and ensure the success of a metrics program. This section summarizes their opinions and will be the basis for the analysis in subsequent sections.

We define a metric as good if it can be used as part of a metrics program to encourage behavior that improves the software development process or product. A metrics program is successful if:

- the metrics collected are used actively and effectively for process and product improvement, and
- the program is long-lasting (that is, it has a high probability of surviving more than two years).

These definitions emphasize that the purpose of collecting metrics and having metrics programs is improvement, not measurement for measurement's sake. These definitions are also broad enough to make them applicable to IT metrics in general (and IT value metrics in particular), independent of whether the software for the IT system is developed from scratch, purchased off the shelf, or customized by an integrator.

Characteristics of Good Metrics

Good metrics have three fundamental characteristics, according to our review of the following sources:
[Armitage and Atkinson 1990] [Jones 1991]
[Basili 1992] [Jones 1993]
[Basili and Weiss 1984] [Kaydos 1991]
[Brown 1993] [Kearney et al. 1986]
[Cusumano 1992] [Pfleeger 1993]
[Daskalantonakis 1992] [Pfleeger and McGowan 1993]
[Grady and Caswell 1987] [Roche 1994]
[Harmon 1992] [Rombach 1990]
[Hronec 1993] [Rubin 1993]
[Humphrey 1989] [Udo 1993]
[Joiner 1992] [Xerox 1991]

These sources describe, in part, the experiences of Hewlett-Packard, Xerox, Motorola, Contel, Hughes Aircraft, NEC, NASA, and others in trying to define good metrics.

Good metrics are:

1. accurate
2. informative
3. easy to obtain.

Each of these characteristics can be expanded into greater detail, as shown in Table 1.
Table 1: Characteristics of Good Metrics

1. **Accurate**
   For a metric to be accurate, it must be:
   a. able to accurately measure the target with little random variation or measurement error.
   b. defined clearly and unambiguously.
   c. objective.
   d. hard to manipulate.
   e. successfully field tested.
   f. robust and repeatable (that is, different people measuring the same thing will get the same result).

2. **Informative**
   For a metric to be informative, it must be:
   a. timely.
   b. easy to interpret.
   c. able to help identify sources of problems and suggest solutions.
   d. able to answer a question about progress towards a goal ([Basili 1992]'s goal-question-metric paradigm).
   e. based on a model or hypothesis of that which is being measured.
   f. able to be compared with goals, baselines, benchmarks, and control limits.
   g. appropriate to the process maturity of the organization.\(^3\)

3. **Easy to obtain**
   For a metric to be easy to obtain, it must be:
   a. economical. Specifically, the metric must have:
      i. low collecting and reporting overhead.
      ii. automated data collection, whenever possible.
   b. associated with a well-defined data collection process.

This is a useful definition because it is succinct yet complete. Further, the definition is internally consistent and sensible. The three major attributes plus the components of each can serve as a quick checklist for evaluating whether or not a proposed or existing metric is a good metric.

---

\(^3\) The Carnegie-Mellon Software Engineering Institute's Process Maturity Model and metrics appropriate for each level of the model are described in detail in [Humphrey 1989]. While the SEI Process Maturity Model emphasizes metrics only for Level 4, the Managed Process, and above, [Jones 1993] and [Pfleeger and McGowan 1993] argue persuasively that metrics are important for all levels; it is just a matter of finding metrics that are appropriate.
The literature makes clear what happens when an organization uses a metric that does not have all three characteristics:

- If a metric is inaccurate (for example, it has a low signal-to-noise ratio or is easy to inadvertently or maliciously manipulate), the organization will grow to distrust the metric.

- If a metric is not informative (for example, if it cannot be compared against a baseline or benchmark in order to establish a context for understanding or if a metric measures something that is unimportant in the eyes of the firm), the organization will become skeptical of the metric.

- If a metric is not easy to obtain (for example, if it must be tracked extensively by hand or requires a data collection process outside of the bounds of normal workflow), the organization will grow tired of the metric.

The result, whether from distrust, skepticism, or fatigue, is disuse. Bad metrics will be abandoned.

The literature also notes that just because something *can* be measured with a good metric does not mean that it *should* be. The corollary to the refrain "what gets measured is what gets done" is "you measure what you can measure." That is, organizations often assess themselves quantitatively with metrics that are easy to get, not necessarily the ones that are most enlightening. For example, this is particularly true with regards to the software engineering community's use of "lines of code" as an input into many traditional metrics. The number of lines of code is trivial to calculate but it is widely acknowledged to be misleading and potentially counterproductive. For example, if the productivity of programmers is measured by "number of lines of code per person per month," the programmers can easily manipulate their coding practices to make themselves look like stars -- without changing the functionality of the code one whit.
There are alternatives to lines of code, most notably, function points. But lines of code can be counted with cheap software tools in an instant while function points can, for the most part, only be counted by hand today in a tedious, time-consuming manner. Thus, lines of code remains in widespread use despite its flaws. This example should caution managers that just because a metric is accurate, (arguably) informative, and easy to obtain does not necessarily mean that the metric can be used well in the organization. The metric may engender undesirable behavior that may ultimately defeat the goals of developing the metric in the first place.

**Characteristics of Successful Metrics Program**

According to our review of the sources listed before as well as [Fenick 1990], [Grady 1992], [Lanphar 1990], [Rifkin and Cox 1991], and [Yourdon 1992], successful metrics programs are:

1. accepted by all
2. well funded
3. durable
4. tied to change processes
5. non-threatening
6. well sized
7. balanced

Each of these characteristics can be expanded into greater detail, as shown in Table 2.

Unfortunately, this definition of successful metrics programs is not as crisp and consistent as that for good metrics. For example, while many authors strongly argue that a metrics program must be non-threatening to the people involved, software development today is still fundamentally a manual process. So for the program to be tied to change processes, as the authors also argue, it must at some point get to measuring and changing people. That is threatening.
The size, scope, and interactions between the characteristics included in this definition result in huge problems for organizations that attempt to start software metrics programs. [Rubin 1993] provides evidence for this claim, noting that of more than 400 metrics programs established by IT organizations since 1980, fewer than 80 -- less than 20% -- have been successful. Greater attention on creating both good metrics and a receptive environment for their use should help reduce the infant mortality of these programs.

The list makes clear that organizational issues should dominate the concerns of those setting up metrics programs. The first four traits (accepted by all, well funded, durable, and tied to change processes) are almost completely divorced from the contents of the metrics; instead, these traits emphasize that the organization has to be molded first to accept the use of measurements to provide the feedback needed for continuous improvement and then to commit to the use of metrics for the long haul. Commitment is crucial; Roche notes, "Without the 'goodwill' of staff, the most detailed measurement programme will fail" [Roche 1994, p. 82].
<table>
<thead>
<tr>
<th>Table 2: Characteristics of Successful Programs</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.   <strong>Accepted by all</strong></td>
</tr>
<tr>
<td>For a metrics program to be accepted by all, it must be:</td>
</tr>
<tr>
<td>a. established in an environment where people are willing to change.</td>
</tr>
<tr>
<td>b. sold heavily to all involved.</td>
</tr>
<tr>
<td>c. defined with measurement processes and goals that are kept close to software developers.</td>
</tr>
<tr>
<td>d. able to measure things important to everyone: those measuring, being measured, and being informed.</td>
</tr>
<tr>
<td>2.   <strong>Well funded</strong></td>
</tr>
<tr>
<td>For a metrics program to be well funded, it must be:</td>
</tr>
<tr>
<td>a. given sufficient resources (funding, time, enthusiasm, and patience) for the long haul.</td>
</tr>
<tr>
<td>b. established with costs recognized and accepted up-front (a metrics program can cost 5-15% of total development costs).</td>
</tr>
<tr>
<td>3.   <strong>Durable</strong></td>
</tr>
<tr>
<td>For a metrics program to be durable, it must be:</td>
</tr>
<tr>
<td>a. run and used by people trained and experienced with using metrics.</td>
</tr>
<tr>
<td>b. sponsored by executives and middle managers who have appropriate but varying degrees of awareness, buy-in, and ownership.</td>
</tr>
<tr>
<td>c. recognized that often &quot;any attempt to improve ... will always make things worse before they get better&quot; [Kaydos 1991, p. 3].</td>
</tr>
<tr>
<td>4.   <strong>Tied to change processes</strong></td>
</tr>
<tr>
<td>For a metrics program to be tied to change processes, it must be:</td>
</tr>
<tr>
<td>a. part of an organization's Plan-Do-Check-Act (PDCA) cycle.</td>
</tr>
<tr>
<td>b. set up with well-defined reporting plans that define what data is needed by whom, when, why, and in what format.</td>
</tr>
<tr>
<td>c. set up with well-defined action plans that show how the metrics will be used for understanding, evaluation, control, and prediction.</td>
</tr>
<tr>
<td>d. set up with well-defined rewards and punishments to encourage desired behaviors.</td>
</tr>
<tr>
<td>5.   <strong>Non-threatening</strong></td>
</tr>
<tr>
<td>For a metrics program to be non-threatening, it must be:</td>
</tr>
<tr>
<td>a. used to measure the process or product, not people.</td>
</tr>
<tr>
<td>b. used in an environment that distinguishes between public and private data.</td>
</tr>
<tr>
<td>6.   <strong>Well sized</strong></td>
</tr>
<tr>
<td>For a metrics program to be well sized, it must be:</td>
</tr>
<tr>
<td>a. made up of a small set of good metrics.</td>
</tr>
<tr>
<td>b. tailored to individual products or projects (not all products and projects need &quot;best practices&quot; processes and metrics).</td>
</tr>
<tr>
<td>7.   <strong>Balanced</strong></td>
</tr>
<tr>
<td>For a metrics program to be balanced, it must be:</td>
</tr>
<tr>
<td>a. composed of metrics that have robust interactions such that attempts to improve one metric at the expense of something else important will be caught by another metric.</td>
</tr>
<tr>
<td>b. run without an emphasis on any one metric to the exclusion of others.</td>
</tr>
</tbody>
</table>
A recurrent danger noted in the literature is that a small band will "get religion" and will try to drive the use of metrics without bringing the rest of the organization along. As with bad metrics, the rest of the organization can suffer from distrust, skepticism, and fatigue of the band's antics, leading them to actively or passively resist the efforts of the metrics leaders. Over time, this will kill the metrics program.

A useful technique for identifying organizational obstacles to the successful implementation of a metrics program is a force field analysis. In such an analysis, the forces for and against the program are listed for each person (by title or name) in the firm. The task of those pushing the program becomes one of taking advantage of the prevailing forces and overcoming the countervailing ones through a judicious choice of metrics, individually-tailored marketing campaigns, careful diplomacy, and so on. An example of a force field analysis is included as Table 3.

Those leading the use of metrics should also remember that metrics are not always needed. In a review of Japanese software measurement practices, Cusumano notes that "even the elite Japanese programming organizations will select, for business reasons, which projects receive the 'best-practice' processes and which do not" because the overhead of best practices is high and may not be worth the effort [Cusumano 1992, p. 13]. Hewlett-Packard, a leading American user of software metrics, agrees with the Japanese: the choice of software development processes and metrics is a function of the degree of difficulty, visibility, and importance associated with the software development effort (see Figure 2). Metrics users should strive to be confident that the cost of applying the metrics that they have chosen is outweighed by their benefit.

With these caveats in mind, those in industry who understand the characteristics of good metrics and successful metrics programs should have a much greater chance of getting metrics used effectively in their organizations for the long term.
Table 3: Sample Force Field Analysis for a Metrics Program

**Division Manager's View**

<table>
<thead>
<tr>
<th>Forces For the Metrics Program</th>
<th>Forces Against</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Productivity improvements necessary for survival</td>
<td>• Cost</td>
</tr>
<tr>
<td>• Improved management control</td>
<td>• Employees' lack of understanding</td>
</tr>
<tr>
<td>• Early diagnosis of problems</td>
<td>• Difficult of establishing valid measures</td>
</tr>
<tr>
<td>• Able to communicate achievements to management</td>
<td></td>
</tr>
<tr>
<td>• Improved ability to plan</td>
<td></td>
</tr>
<tr>
<td>• Ability to validate staffing levels</td>
<td></td>
</tr>
<tr>
<td>• Ability to recognize good units</td>
<td></td>
</tr>
</tbody>
</table>

**Department Manager or Front-Line Employee's View**

<table>
<thead>
<tr>
<th>Forces For</th>
<th>Forces Against</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Top management wants it</td>
<td>• More reports to complete</td>
</tr>
<tr>
<td></td>
<td>• Closer management control</td>
</tr>
<tr>
<td></td>
<td>• Might be embarrassing</td>
</tr>
<tr>
<td></td>
<td>• Boss doesn't understand our work</td>
</tr>
<tr>
<td></td>
<td>• Nothing in it for me</td>
</tr>
<tr>
<td></td>
<td>• Excuse to cut our resources</td>
</tr>
<tr>
<td></td>
<td>• Can't measure what I do</td>
</tr>
</tbody>
</table>

Source: [Grady and Caswell 1987, p. 93]
Figure 2: Balancing Risk and Overhead in Software Projects

- **High risk area:** too little control & measurement
- **High overhead area:** too much control & measurement

Source: Hewlett-Packard, Corporate Engineering, personal communication, 1994
Section 3: Value Metrics at the Avalon Corporation

The question of how to best measure the value of CIM systems is one that has bedeviled many firms, including the Avalon Corporation. The author spent six months on an internship at Avalon helping the company explore this issue. This section is based on that experience and critiques a CIM value metric proposed by Avalon as well as two other approaches for measuring the value of CIM systems.

The internship was arranged through the MIT Leaders for Manufacturing Program, of which the author is a fellow and Avalon is a sponsor. The author worked on various projects that were led by members of the corporate Manufacturing Systems and Operations Research (MS&OR) department, the Components Operations' (CO's) Manufacturing Planning and Information Systems department, and the Systems department of the CO's Silicon Division. The project most closely tied to the evaluation of a value metric was work for the Plant Floor Action Team that was championed by the heads of MS&OR and the Silicon Division's Systems department. This metric and its history will be discussed first.

Avalon's Systems Efficiency Program (SEP)

The Plant Floor Action Team (PFAT) was one of many teams that compromised Avalon's Systems Efficiency Program (SEP). The SEP was established in the early 1990's with the ambitious goal of doubling the productivity of Avalon's systems departments within five years.

The motivation for Avalon to make such dramatic improvement is clear from the results of IT benchmarking conducted against its major Japanese competitors. While Avalon's IT spending as a percentage of revenue is in line with other large American manufacturing firms, it is substantially higher than that of its Japanese
competitors. IT spending per unit of product sold is also commensurately higher, as is the number of IT employees [Avalon 1992a].

Much of the Japanese advantage seems to come not from more productive IT investment, but from simpler business practices, simpler organizations, and tighter cost controls [Avalon 1992a]. By simplifying processes and reducing the number of exceptions therein, the Japanese firms find that they can automate more cheaply or avoid automation all together. This is particularly true at the largest of the Japanese competitors, a firm that is often the leading benchmark for quality, cost, and performance throughout Avalon.

While Avalon is moving towards constantly simplifying its processes to better compete with the Japanese, it believes that its reengineering efforts will be made possible only through the use of IT. Thus, over time, Avalon expects to use more IT, not less. It thinks that its growing skills in effectively developing and deploying IT will provide it with an advantage over its Japanese rivals.

As such, the Systems Efficiency Program plans to meet its goal of doubling productivity by doubling the amount of output and holding headcount constant. This stands in contrast to a similar productivity improvement program underway in Avalon's finance departments. There, the volume of work will be held constant; headcount will be halved. If the SEP is unable to demonstrate that it has met its goal of doubling productivity by doubling output, the fate of the finance departments may become the fate of the systems departments -- providing a strong motivation for the success of the SEP.

**SEP's Plant Floor Action Team (PFAT)**

It was in this context that the PFAT was formed. Its charter was to find ways of doubling the productivity of the development of CIM systems within five years. PFAT set up a large number of subteams that developed software development process
improvements (e.g., rules for effective rapid prototyping, factory-to-factory replication of CIM systems, remote data collection in factories, etc.), organizational improvements (e.g., a shared resource center to bring together heretofore scattered systems engineers and technologies), and more. These developments were quickly and successfully deployed.

PFAT also established a subteam to create metrics to monitor its progress in improving the development of CIM systems. In 1992, nine metrics in the following five categories were proposed:

1. Effectiveness
2. Delivery
3. Efficiency
4. Quality

PFAT hoped to be able to show that the effectiveness and efficiency of the production of CIM systems was doubling while on-time delivery, quality (as measured by post-launch defects), and customer satisfaction were improving.

**PFAT's Systems Request Effectiveness (SRE) Metric**

The effectiveness metrics are the ones most germane to a discussion of value. While three effectiveness metrics were defined, the last two are simply variations on the first so we will concentrate only on that one: the Systems Request Effectiveness (SRE) metric.

SRE was defined as follows:

\[ SRE = \frac{\text{net present value of customer value in dollar terms}}{\text{project work hours}} \]

The SRE for a project is supposed to be measured twice: an estimate is to be prepared once when the project is approved and an actual value is to be prepared after the system
has been deployed. (In this way, projected versus actual results can be compared.) Work hours are tracked in a corporate database. If all or part of a CIM system is purchased from a third party, that cost is divided by a standard rate to get an equivalent number of work hours. The net present value is calculated using Avalon's annual discount rate.

Customer value is calculated as the sum of the hard and soft dollar savings on which the customer has agreed. PFAT expected that such data would be drawn from standard financial documents and project justifications.

All of the data needed to calculate the SRE metric are to be reported by CIM project managers at quarterly intervals to PFAT. PFAT in turn reports summaries of all of its metrics to the SEP.

During early 1993, PFAT attempted to establish baselines from all of its metrics, including SRE, by gathering data on existing CIM systems from CIM project managers. Unfortunately, of the 33 projects that reported data, only six included a dollar figure for customer value. That fact, along with systems managers' complaints over the worth of the SRE metric, led PFAT to ask the author to help study the PFAT metrics, particularly the SRE metric.

To do so, a dozen structured interviews were conducted by the author and another Avalon employee with managers of plant-floor systems development in Avalon's Assembly Operations, Fabrication Operations, and the five divisions of the Components Operations. The interviews revealed the following with regards to the Systems Request Effectiveness metric:

1. Less than half of the managers could provide any customer value data for the SRE metric, and of the ones who could, many considered the data to be highly suspect. While PFAT expected that the data could be drawn from standard financial documents, this simply was not the case.
2. There was consensus that it is often very difficult to determine customer value in dollar terms with any exactitude either before a project is started or after it is launched.

3. Dollar amounts of value are not always used to determine project priorities.

4. Whether or not the projected customer value was realized is seldom verified.

Clearly this metric was in trouble. While potentially informative, the interviews revealed that the metric was neither accurate nor easy to obtain: it was not a good metric.

Yet many Avalon managers wished that they could measure the value of their CIM systems in dollar terms. Being able to do so would make it easier to pick the most profitable projects and to persuade plant and division managers that they were doing the right thing. Therefore, the thesis will now explore why these CIM managers were having such difficulty in translating value into dollar terms and to try to discover if there were solutions to these problems.

**Dollar Value, a Poor Metric for Many CIM Projects**

Information drawn from these interviews as well as financial planning and project justification documents from a variety of CIM efforts at Avalon leads to two conclusions. First, CIM projects generate three different types of savings:

1. direct cost savings ("hard" savings)
2. indirect cost savings ("soft" savings)
3. opportunity cost savings.

The nature of these savings is explained in detail below but in general, direct cost savings are the ones that can be most readily measured in dollar terms while such measurements are much harder to make for indirect and opportunity cost savings. Second, CIM projects tend to fall into one of four categories:

1. stand-alone projects
2. new product launches
3. compliance projects
4. infrastructure projects.

Easily-measurable direct cost savings tend to be associated only with stand-alone projects. Indirect and opportunity cost savings generally dominate the value derived from the other three types of projects. Even stand-alone projects will have these other sorts of savings, often to the point that these hard-to-measure savings predominate. To understand why all of this is so, the nature of these three types of savings and four types of projects will be examined in more detail.

**Cost Savings**

As indicated in Table 4, direct cost savings are easy to calculate in dollar terms because they are tracked in detail by traditional cost accounting systems.\(^4\) Not so for most indirect cost savings. The immediate costs and benefits of improvements in flexibility, speed, or learning are not tracked in a ledger and are not readily apparent in a balance sheet, income statement, or cash flow statement.

This does not mean that it is impossible to calculate the value of an indirect cost savings in dollar terms. For example, an improvement in quality can lead to a reduction in warranty costs, a large proportion of too many plant managers' budgets. But it may take a long time for this savings to be realized; at Avalon, assuming a three year product development cycle, a five year model life, and a three year warranty, it may take up to eleven years from the start of the project where the dollar amount of CIM-related warranty reductions was estimated to see if those savings were fully achieved. Direct cost savings tend to be more immediate; install a new CIM system,

\(^4\) Note that the use of activity-based costing (ABC) systems instead of traditional labor-and-overhead cost accounting systems may make it much easier to track the effects of CIM systems on a plant's cost drivers. But these effects will still tend to be direct cost savings, not indirect or opportunity cost savings.
ramp up the learning curve, and watch labor, overtime, inventory, or scrap levels fall right away. Being able to verify the accuracy of an estimate early makes it easier to hold responsible those making the estimates. Presumably this helps to encourage good estimates.

<table>
<thead>
<tr>
<th>Table 4: Examples of Cost Savings</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Typical Direct Cost Savings</strong></td>
</tr>
<tr>
<td>Savings are reductions in:</td>
</tr>
<tr>
<td>Direct labor hours</td>
</tr>
<tr>
<td>Overtime</td>
</tr>
<tr>
<td>Defects and scrap</td>
</tr>
<tr>
<td>Inventory</td>
</tr>
<tr>
<td>Inventory underutilization</td>
</tr>
<tr>
<td>Per-unit overhead allocations</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Typical Indirect Costing Savings</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Savings are increases in:</td>
</tr>
<tr>
<td>Quality</td>
</tr>
<tr>
<td>Flexibility</td>
</tr>
<tr>
<td>Speed</td>
</tr>
<tr>
<td>Learning</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

An example of a successful CIM project that generated only indirect savings is a paging system installed in one of Avalon's factories. For a small investment (tens of thousands of dollars), a system was developed that allowed plant employees as well as the plant's automated statistical process control and maintenance management systems to send alphanumeric pages to employees equipped with personal pagers. This system allowed critical information to flow virtually instantaneously to key people no matter where they were located in the plant, freeing people to roam more loosely in the plant.
while still being able to respond quickly to problems. The plant's former systems manager argues that this CIM system was a huge success because it sped up problem solving and changed work relationships for the better.

But measuring the value of this success in dollar terms remains problematic. There was great skepticism before the project was approved that the investment would provide an adequate return because translating "speedy problem solving" and "better work relationships" into a number that the accountants can use is difficult. Interestingly, even though now the plant management swears that the paging system is vital to its operations, making the case that the system creates value remains difficult. Only three other plants out of dozens of candidates have copied the system because, among other things, making the financial case for these indirect cost savings is not readily doable.

The systems manager who oversaw the paging project complained about Avalon's inability to spread the good word about successful systems. At a recent worldwide conference for Avalon's systems managers, he saw "fifty examples of great systems that only existed in one plant." If it were easier to turn these anecdotes into numbers, he argued, they would be replicated quickly.

Sometimes the benefits of indirect cost savings can be measured in dollar terms but only through the use of long inference chains or stochastic models. For example, one benefit of improved quality is higher profits. As Figure 3 shows, improving quality can make customers happier, increasing the likelihood that they and their friends will buy Avalon products, thereby boosting Avalon's bottom line. But quantifying the exact (or even approximate) effects of an increase in the level of quality on the other levels in the system shown in Figure 3 is difficult, making it tough to project the return to a quality-improving investment in CIM. Even if such a projection were made, so many other big factors not included in this simple model affect customer
satisfaction, sales, and profitability that correlating after the fact a change in profitability with an improvement in quality is fraught with danger.

Figure 3: Quality's Relationship to Profitability

It is possible to build sophisticated economic models to capture the value generated by these soft savings. While a specific example from Avalon cannot be cited, a general review of the literature indicates that model building runs into the following types of problems:
1. Construction of sophisticated models to measure the dollar value of CIM systems is possible but the number of assumptions required to build the models can keep them from being widely accepted in the organization. Having a good model alone does not guarantee acceptance.

2. The assumptions also lead to probabilistic conclusions. While understanding that there can be a range of outcomes is important, it makes it more difficult to extract a single number that represents the value of a CIM project and plug it into a metric.\(^5\)

3. Politics enters into decision making. Since models are based on the assumptions of their creators, the resulting numbers can reflect political reality more than economic reality, again potentially making the numbers inappropriate for use in good metrics.

In sum, while models can be useful for decision making, models of value in dollar terms tend to produce numbers not fit for use in good metrics.\(^6\)

---

\(^5\) Given that soft savings are so hard to quantify exactly, one might expect to find most firms using some type of sensitivity analysis in their calculations. However, the study described in [Sohal et al. 1990] of 61 Australian manufacturing firms found that 40% of the firms did not conduct sensitivity analysis when planning investments in advanced manufacturing technology.

\(^6\) Because the author was unable to find an example of sophisticated model building within Avalon, to better understand this point, consider the following case that is drawn from the literature. The Burlington Northern railroad was considering implementing a sophisticated railroad management system called ARES. To measure indirect cost savings (e.g., "how much more would a customer be willing to pay for a 1% improvement in service?" [Kaplan and Hertenstein 1991, p. 10]), the firm:

- built a demand elasticity model using conjoint analysis,
- ran stochastic simulations, and
- constructed other models using Markov analysis.

The output was subjected to sensitivity analysis, allowing the systems department to construct a graph showing the probabilities of achieving a range of NPV targets.

Unfortunately, Burlington Northern's management did not buy the results of the model:

Executives ... did not fully believe all aspects of the ARES benefits analysis, particularly the service-price elasticities. The marketing department considered the estimates overstated and wondered why its people had not been more involved in developing this analysis beyond providing suggestions on market research firms, research sites and questionnaire content. [Kaplan and Hertenstein 1991b, p. 2]

The marketing department had estimated that a 1% improvement in service reliability would lead to a 0-0.3% price gain. Two outside analyses predicted 0.2-3.3% and 2.8-6.5%, respectively. The systems department decided to use a nominal value of 2%, leaving it open to criticism from its marketing counterparts. Thus, the numbers derived from the model were not accurate, per the definition of good metrics because they were unable to measure the target with little measurement error, were subjective, and were easy to manipulate.
Many of the problems with indirect cost savings are also true for opportunity cost savings. An opportunity cost savings is one where a potential cost is avoided. For example, if a plant were running at effective capacity and a CIM system that increased effective capacity by 25% were to be installed, the CIM system has let the plant avoid the cost of building new capacity or the cost of missing out on potential orders if demand were to rise. The CIM system has created an opportunity cost savings relative to the next best alternative.

Measuring the magnitude of opportunity cost savings is difficult. In the case of a plant-capacity-boosting CIM system, the firm has the opportunity to buy an option on meeting future higher demand. Options Pricing Theory provides techniques for quantifying the dollar value of such options but applying it here is troublesome. Because the markets for the items used as inputs into the equations are often illiquid and not traded on the open market, it is hard to establish accurate numbers to plug into the pricing model.

These problems in easily quantifying the amounts of indirect and opportunity cost savings lead project managers and financial analysts to undervalue their importance in their cost-benefit analyses. As Kaplan notes, "[c]onservative accountants who assign zero values to many intangible benefits prefer being precisely wrong to being vaguely right" [Kaplan 1986, p. 92].

This emphasis on exactitude at the expense of accuracy is dangerous because it allows CIM project managers to avoid having to even think about putting rough numbers on their claims. At Avalon, in many of the CIM project proposals that were reviewed by the author, the main value of the project asserted in the text of the project proposal was not included in the financial calculations. For example, a new

(Note that even gaining interdepartmental consensus is no guarantee over overall acceptance. In another case, a team came to agreement on the cost savings expected from an IT project, only to have that number "reduced by the ... division manager to ensure that the projections were realistic" [Clark and McKenney 1990, p. 7].)
maintenance management system was approved for a CO plant because it would improve product quality and allow better control over preventative maintenance in the plant, something very important to the plant as it sought to win Avalon's new Preventative Maintenance Excellence Award [Avalon 1993b]. The value of the project was clear to plant management because, in a tightly capital constrained environment, they approved the project only after canceling another high-priority one. But in the financial calculations, the only benefit ascribed to the CIM system was a direct cost savings -- an end to a contract for outsourcing preventative maintenance systems service. While this one savings was enough to generate a time-adjusted rate of return greater than Avalon's very high hurdle rate and to achieve a rapid payback period, the stated dollar value of the project had nothing to do with what its backers thought was most important about the project.7

In summary, CIM project managers report that while they can get hard figures for direct cost savings, they have a rough time measuring accurately and easily the value of indirect and opportunity cost savings in dollar terms. Direct cost savings can be measured easily because they generally affect line items in existing plant budgets. Indirect cost savings have more distant effects on the plant or firm's bottom line that are difficult to measure ahead of time and to verify after the fact. Opportunity cost savings force a comparison against the next best alternative, often permitting only a rough estimate of their relative value.

---

7 Interestingly, while the financial calculations showed all of the initial deployment costs and on-going savings, they did not show any new on-going costs (e.g., in-house Avalon personnel devoted to maintaining and enhancing the new system). If these "hidden costs" (buried in someone else's budget) had been recognized, the plan may not have cleared its hurdles.
Types of CIM Projects

As noted earlier, these hard-to-quantify indirect and opportunity savings predominate in three of the four types of systems projects. All four types are detailed here:

1. **Stand-alone projects** are ones where a CIM system is added to an existing manufacturing line. Direct cost savings -- and some indirect cost savings -- are generally easy to measure here by comparing the difference in costs before and after the system is installed. An example of a stand-alone project at Avalon would be one that was proposed to "improve efficiency and utilization of the Schuler Press at [an Assembly stamping plant] by developing a system which will reduce press and automation repair time" [Avalon 1993a, p. 1]. The pager system described earlier is also an example of a stand-alone project.

With a CIM system being added to an existing line for which financial and process performance data already exist, seeing the projected improvement materialize or not should be easy. For that reason -- and to keep its plant managers honest -- Avalon requires strict accountability when plants sign up for stand-alone projects. If a plant agrees that its raw materials costs will be reduced by a new system, its raw materials budget will be cut accordingly by the division finance staff. If headcount is to be reduced, the plant will have to "give the social security numbers of the people to be fired to the finance guys," noted one Avalon manager. This avoids the problem faced by another firm:

Most participants believed that the savings would eventually offset the cost, but virtually everyone believed the savings in labor would come from someone else's area. [Clark and McKenney 1990, p. 8]

One other aid to accurate estimates of direct cost savings is replication. Avalon promotes the replication of identical systems in many different plants as a way of increasing the productivity of the systems department and of spreading useful
systems rapidly. When systems are replicated, plant and division managers want to know how the first system that was installed did versus its original estimate. The CIM managers' knowledge that their projections will be verified is an incentive for them to make those projections accurate. Also, the actual data available after the first installation provides concrete numbers for improving future projections.

Altogether, for these reasons, Avalon CIM managers find the direct cost savings calculated for stand-alone projects to be reasonably accurate.

2. *New product launches* are projects where a CIM system is being included as part of an entirely new manufacturing line or factory. Here, while CIM systems are vital, alone they create no value. It is the union of all the functional areas in the plant that produce value, as shown in Figure 4. The marginal benefit in value to the customer derived by a marginal increase in systems spending is hard to estimate in this complex system.

![Figure 4: How Value Is Created in New Product Launches](image)

Managers in charge of new product launches are willing to pay for CIM systems because they recognize the indirect cost savings (in terms of improved quality, flexibility, speed, etc.) and opportunity cost savings (e.g., labor savings versus a
hypothetical plant without CIM) that these systems can provide. But these managers do not necessarily require that these savings be quantified.

At Avalon, the funding for CIM systems in new product launches is generally included as a small line-item expense (perhaps 1-3% of total costs) in the launch budget. CIM project managers try to maximize the utility of the systems that they choose to implement within that constraint but they are not required to make any formal dollar-wise calculation of the benefits of CIM. This is especially true in the Assembly Operations and Fabrication Operations where they have large numbers of plants making similar products, allowing for a high commonality of systems across plants (unlike the Components Operations with its high variety of products and plant processes). Assembly and Fabrication CIM managers have created suites of applications that they roll out into all new plants and lines. While they try to drive the costs of replication down, they do not have to justify the deployment decision in financial terms.

3. **Compliance projects** are ones where a CIM system is needed to comply with a regulatory or contractual requirement. Examples cited during the interviews with Avalon managers are listed in Table 5.

<table>
<thead>
<tr>
<th>Category</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Safety regulations</td>
<td>Tracking safety device serial numbers with product identification numbers to satisfy federal safety requirements</td>
</tr>
<tr>
<td>Environmental regulations</td>
<td>Monitoring emissions from paint operations in assembly plants to satisfy Environmental Protection Agency requirements</td>
</tr>
<tr>
<td>Import/export regulations</td>
<td>Tracking U.S.-built products shipped to Mexico to satisfy Customs requirements</td>
</tr>
<tr>
<td>Union agreements</td>
<td>Tracking skilled labor overtime allocation to satisfy union agreements</td>
</tr>
</tbody>
</table>
These systems are approved not because they create any additional value for the business but because they preserve value generated elsewhere; as such, they are an opportunity cost savings. They are the cost of remaining in business and, as such, plant and division managers attempt to hold down their cost to the minimum required to satisfy the regulation or contract. No detailed financial calculations are needed because the management knows that the cost of compliance is far less than the present value of the profits that will be generated by the operation and preserved by the CIM system. Because the financial calculations will not be an aid in decision making, there is no need to make them.

4. **Infrastructure projects** are ones where computing infrastructure (networks, servers, database systems, etc.) are installed in factories. Like a road with no cars, infrastructure alone generates no value. It is the people and the software that use the infrastructure which produce value.

The first system to need to use the infrastructure is often forced to bear the burden of the entire cost of the infrastructure. This increases the cost/benefit ratio of the first project (sometimes to the point where the costs exceed the benefits) while lowering that of later projects which get a free ride. (Note, however, that once the infrastructure is installed, it should be treated as a sunk cost and not charged against any future CIM projects.)

The difficulty in calculating the value that can be derived from a well-utilized CIM infrastructure drives many organizations towards treating infrastructure as a compliance project. For example, Avalon's Components Operations has a standard hardware, networking, and database architecture that it requires plants to have installed before they can use CO-produced CIM systems.

In sum, this view of the nature of CIM projects finds that:
• Stand-alone projects are the only ones to predominantly have direct cost savings that are easy and accurate to measure. But stand-alone projects often also have hard-to-measure indirect and opportunity cost savings.

• New product launches have indirect and opportunity cost savings but the magnitude of these benefits is not calculated as an input into decision making.

• Compliance projects are treated as a necessary cost of doing business. Financial analysis focuses only on minimizing the costs of such projects, not calculating the value preserved in opportunity cost terms.

• Infrastructure projects have no value on their own and are generally funded by burying their cost in a stand-alone project or new product launch or by treating them as a compliance project.

Therefore, only for stand-alone projects which derive most of their value from direct cost savings would we expect CIM project managers to have the numbers needed to calculate PFAT’s Systems Request Effectiveness metric. For other types of projects, the numbers simply do not exist; pushing managers to provide them will lead only to responses of self-serving guesses or to no responses at all.

**Continued Use of the SRE Metric**

If getting good numbers to plug into the SRE was so difficult, why did PFAT persist in trying to use the metric? It was clear that use of the metric was not going well: as noted earlier, PFAT had run into problems when it tried to gather data to establish a baseline for the SRE metric. Less than one-fifth of the CIM projects that reported metrics data to PFAT included the savings data needed to calculate the SRE metric. PFAT had cajoled the CIM project managers and their supervisors over a long period of time to get the necessary numbers but this effort was not fruitful. These problems suggested that CIM managers at Avalon were actively and passively resisting providing data for this metric and that, given the preceding analysis' conclusion that the
SRE metric fails the test of a good metric (see Table 6), the SRE metric should have been abandoned. Yet PFAT continued to advocate the use of the SRE metric.

There are three reasons that PFAT persisted in pushing the metric:

1. PFAT members thought the necessary data was available in standard Avalon financial forms required for project approval.

2. PFAT members believed that it was important not to get caught in an unending debate over metrics definitions. Instead, they wanted to move on to active use of albeit imperfect metrics in order to get the job done.

3. Avalon's "good news culture" encouraged PFAT to hide the problems encountered with the SRE metric from upper management.

Each of these reasons is described in more detail below.
Table 6: Evaluation of the Systems Request Effectiveness Metric

1. **Accurate**
   For a metric to be accurate, it must be:
   - a. able to accurately measure the target with little random variation or measurement error.
   - b. defined clearly and unambiguously.
   - c. objective.
   - d. hard to manipulate.
   - e. successfully field tested.
   - f. robust and repeatable (that is, different people measuring the same thing will get the same result).

2. **Informative**
   For a metric to be informative, it must be:
   - a. timely.
   - b. easy to interpret.
   - c. able to help identify sources of problems and suggest solutions.
   - d. able to answer a question about progress towards a goal (‘Basili 1992’ s goal-question-metric paradigm).
   - e. based on a model or hypothesis of that which is being measured.
   - f. able to be compared with goals, baselines, benchmarks, and control limits.
   - g. appropriate to the process maturity of the organization.

3. **Easy to obtain**
   For a metric to be easy to obtain, it must be:
   - a. economical. Specifically, the metric must have:
     - i. low collecting and reporting overhead.
     - ii. automated data collection, whenever possible.
   - b. associated with a well-defined data collection process.

First, having developed CIM systems throughout Avalon, PFAT members had all encountered the standard financial forms required for project justification. The SRE metric had been designed with the assumption that most, if not all, CIM projects filled out these forms and that therefore, the data for the SRE metric could be drawn from them. PFAT was aware of the difficulties in quantifying indirect and opportunity cost
savings but thought that the direct cost savings and reasonable estimates of the other types of savings would be included on these two forms.

This belief was wrong for two reasons. One, as the interviews conducted by the author revealed, CIM projects tend to fall into one of the four categories discussed earlier and the financial documents PFAT wanted to examine are generally used in ways appropriate to PFAT's needs only for stand-alone projects. For new product launches where CIM is a line item in the overall budget, no monetary cost-benefit analysis was required; the same was true for many compliance and infrastructure projects. Second, as noted before, when the financial forms were filled out, they were often filled out just enough to ensure that Avalon's return-on-investment hurdle was surpassed and not necessarily to include estimates in dollar terms of all of the sources of value within the project. The finance department in Avalon historically has been very strong and very willing to hold plants to their financial commitments (whether or not the CIM systems actually produced the improvements they promised to). This has made the plants quite shy of signing up financially for any of the CIM systems' benefits beyond the minimum required to get the needed signatures.

Thus, what the interviews showed was that one of PFAT's underlying assumptions about how the metric would be gathered was wrong. In spite of this evidence, there was still an undercurrent of feeling within PFAT that the data should be available (which can be taken as either a practical or moral/Puritan view) and that, therefore, the SRE metric should still be used.

The second reason that PFAT persisted with the SRE metric was the recognition that no metric is perfect. PFAT understood what many of the sources cited in Section 2 noted: all metrics have flaws and while a sustained debate over a metric's definition
may correct serious flaws, it cannot fix them all. Therefore, as the utility of a metric is derived only from its use, PFAT was willing to accept what it thought were minor imperfections in the metric's definition in order to avoid an endless cycle of debate and modification. PFAT wanted to get on with the business of using the SRE metric and demonstrating the value CIM systems were creating for Avalon.

This is an eminently reasonable view. There were quarrels on the part of PFAT and the interviewed CIM managers over all of the PFAT metrics. Problems were being encountered as PFAT tried to ramp up the use of all nine metrics in each of the five metrics categories. Before a new metric can be used by management for understanding and improvement, it must move sequentially through seven stages:

1. **Defined:** The general scope of the metric and the goal and question to which it is related must be defined. The exact definition of the metric must be written down.
2. **Distributed:** The definition must be sent to those affected by the metric (in terms of data gathering and use of the results).
3. **Accepted:** Those to whom the metric definition was distributed must agree that the metric's definition and purpose are sensible and useful.
4. **Collected:** All of the data necessary to calculate the metric must be collected.
5. **Calculated:** The metric must be calculated from the collected data.
6. **Returned:** The results of the metric calculations must be returned to an appropriate audience.
7. **Used:** The returned metrics must be used actively by the audience for process improvement.

The interviews revealed that all five categories of PFAT metrics were at different stages of this progression. Because none of the metrics had yet reached the final,

---

8 Describing Motorola's software metrics program, Daskalantonakis notes that it is "better to start from a set of metrics addressing important improvement areas, and evolve these metrics over time, instead of debating forever, trying to find perfect metrics" [Daskalantonakis 1992, p. 1008].
"used" stage, there were issues about all of the metrics with regard to how good they were or how they should be used. On the surface, the controversy over the SRE metric seemed no different than that over any of the other metrics.

But Figure 5 reveals that there was a major difference. The delivery, efficiency, quality, and customer satisfaction metrics categories were all accepted by the bulk of Avalon CIM managers who would gather and use the metrics. But those same people resoundingly rejected the effectiveness metrics.

As described earlier, the CIM managers rejected the SRE metric because they could not easily provide the needed data and because they were justly suspicious that the metric would ever be able to capture the true value of their work. Some of the managers also said that they thought that the metric was being used more to prove a point with the Systems Efficiency Program than to help them improve their day-to-day CIM project management practices. Since the metric was not useful to them, they saw no reason to bother with it, either in terms of supplying data to calculate the metric or help PFAT to improve the definition.

PFAT, in its drive to start using the imperfect metrics, had failed to recognize that the effectiveness metrics most likely would never be able to pass through the "accepted" stage to get to the "used" stage. A reasonable willingness to accept minor flaws led them to fail to recognize the serious flaws in the SRE metric.

The third reason for PFAT's persistence over the SRE metric suggested why PFAT held onto these first two reasons in spite of the evidence against them. Avalon has a "good news culture," one that encourages employees to report only good news upwards to management. One of the systems project managers interviewed by the

---

9 Note that all of the metrics had been defined and distributed except for the customer satisfaction metric. The interviews found that everyone accepted the notion of a customer satisfaction metric and there was general agreement at a broad level what a customer satisfaction rating metric should look like. But because so many different parts of Avalon already had CIM customer satisfaction surveys with slightly varying questions, it was recognized that PFAT's definition of the metric would not work in all circumstances so there was a move within PFAT to redefine the metric's definition somewhat.
author explained that Avalon's history as a strongly hierarchical, centralized, Rust Belt company had given rise to this culture. The mark of an excellent Avalon manager, he said, was the ability to listen to a subordinate detail the 27 excellent things he or she had done in the past week and to then ferret out the 28th thing that was "really, really bad."

PFAT's SRE metric and other similar metrics produced by other SEP teams were clearly very important to the Systems Efficiency Program and to its head, Avalon's Chief Information Officer. It was these value metrics that would allow the SEP and CIO to proclaim success in meeting SEP's goals of doubling productivity with a flat headcount.

While PFAT was given the responsibility for gathering the effectiveness metrics data and for implementing programs to train CIM teams on how to improve their productivity, it was not given any authority over those CIM teams to enforce productivity improvements. Nonetheless, should the SRE metric fall short of its goal, PFAT would most likely be held at least partially responsible for that major failure.

But it would be a while before that problem would arise. In the meantime, PFAT had to face mounting criticism of its SRE metric. For PFAT to have rejected its plan for measuring the value of CIM, six months to a year after the plan had been promulgated, would have been problematic. Simply rejecting the SRE metric and offering no alternative would not be acceptable; PFAT would need to offer up an alternate way of measuring value. As the discussion in this thesis has indicated, there do not appear to be generalized measures of the value of IT in widespread use beyond customer satisfaction ratings and clearly inadequate financial calculations. PFAT was in a bind.

The incentives here were aligned well with Avalon's good news culture. The environment in general encouraged PFAT to hide its problems; the lack of alternatives to the metric reinforced this. PFAT responded to this prodding by doing what in many
ways was the most sensible thing for PFAT members to do: continuing to push the use of the metric.
Figure 5: State of the PFAT Metrics Roll-Out

Effectiveness

Defined  Distributed  Accepted  Collected  Calculated  Returned  Used

Delivery

Defined  Distributed  Accepted  Collected  Calculated  Returned  Used

Efficiency

Defined  Distributed  Accepted  Collected  Calculated  Returned  Used

Quality

Defined  Distributed  Accepted  Collected  Calculated  Returned  Used

Customer Satisfaction

Defined  Distributed  Accepted  Collected  Calculated  Returned  Used
The March of Folly

The historian Barbara Tuchman would describe this as a march of folly. The SRE metric was recognizably flawed yet it continued to be used. The author believes that over time, these flaws will prove fatal to the metric. Like so many other bad metrics, it will fall into disuse and disrespect and eventually will be quietly abandoned.

This conclusion is not meant to criticize the Plant Floor Action Team. For a profit-maximizing firm, it is sensible to try to measure the value of CIM systems in dollar terms, to try to understand how CIM improves the bottom line. Financial calculations are required for many projects in Avalon so it was reasonable to try to add those project-level numbers together to create an overall sum of CIM's contribution to the business. Dollar amounts are comparable and aggregatable across disparate CIM projects so there was a strong motivation to try to capture value in dollar terms.

For these reasons, PFAT proposed to measure value with the SRE metric. But this metric was not working out to plan. The difficult-to-quantify nature of many of the savings generated by CIM systems and the varying nature of the types of CIM projects meant that the numbers needed to calculate the metric were not available. Specific circumstances and the general culture at Avalon conspired to drive PFAT to continue to try to use the SRE metric even when its deficiencies were apparent.

This is not to argue that financial analyses of CIM systems are unimportant; justifying them on "faith alone" is foolhardy, as [Kaplan 1986] emphasizes. Developing numbers that provide insight is important for both systems and plant managers. When used well, even inexact figures can be a valuable input into decision making. But this inexactitude makes them vulnerable to corruption if they are used to measure the very people who report them. Beyond ones that measure direct cost
savings found in stand-alone projects, metrics for calculating the value of CIM in dollar terms will be bad metrics. They should not be used.

**Other Attempts to Measure Value in Dollar Terms**

To reinforce this point, two other attempts at Avalon to measure the value of CIM systems in dollar terms are discussed here. One was a study undertaken in Avalon's Music Division and the other was a table put together by the Plant Floor Action Team. Both provided some interesting results but both also depended on fundamentally inaccurate or incomplete dollar figures.

The Music Division's study, detailed in [Maisonville 1993], was designed to determine what benefits the division derived from CIM. The study found that together, four of the division's five plants saved millions of dollars annually because of their CIM systems. A survey of plant personnel conducted for the study also found that plant employees were strongly and universally persuaded that CIM systems:

- are important to operations
- are an integral part of the manufacturing process
- have decreased mistakes and improved quality.

The survey also found substantially less agreement that CIM systems have:

- lowered overall operating costs
- improved manufacturing cycle time
- increased flexibility of manufacturing lines.

To estimate the magnitude of CIM-generated cost savings, an opportunity cost approach was used. Plant employees were asked to estimate how much more it would cost to run the operation if there were no CIM systems. The primary increase in cost would be due to an increase in the number of direct labor hours as people were substituted for computers.

There are three problems with the approach used in this study:
1. **Inaccurate**: While the plants probably should know better than anyone else the cost savings generated by the equipment in the plant and while looking at the cost of the next best alternative is a sensible valuation technique, the numbers generated by the plants are merely guesses. No methodology was defined to ensure any consistency or completeness in the calculations. Thus, different people within the same plant were able to come up with different results. For these reasons, the multi-million dollar savings included in the report is, at best, simply an order-of-magnitude approximation of some of the benefits of CIM systems.

2. **Incomplete**: The study's introduction noted:

   The only benefit quantified was that of cost. Although benefits other than cost may in fact be far more significant, no estimates were made on the effects of CIM systems on customer satisfaction, quality, timing, flexibility, contractual requirements, empowerment, etc. [Maisonville 1993, p. 1]

   Because these "soft" benefits were hard to quantify in dollar terms, they were ignored, making the overall calculation of value incomplete. The study also ignores the **cost** of providing CIM systems to the division (which the study's author, in an interview with the thesis's author, estimated to be about the size of the savings included in the study).

3. **Contradictory**: While the study claims that the annual cost savings attributable to CIM are large, it is not clear that the plants agree. Of the 109 people who replied to the survey question asking if CIM systems have lowered overall operating costs, only 49% agreed or strongly agreed. The rest either disagreed or neither agreed or disagreed. The study does not address this conflict between its conclusions and the plant's opinions.

   These problems mean that the study's conclusions should be approached with caution. Like the PFAT's SRE metric, the study appears to have stumbled over the fact that accurate and easy-to-obtain measures of the dollar value of CIM are not available.
The table prepared by PFAT tripped over this fact as well. Just like the Music Division survey, PFAT was trying to build an overall estimate of the value of CIM, this time for Avalon's Fabrication Operations (FO). As shown in Table 7, the table was prepared as follows:

- The rows were formed by the various CIM systems employed in FO.
- The columns were formed by the 14 plants in FO.
- Each cell in the resulting table was filled in with the dollar value of the CIM system in that plant, if that system had been installed.

To determine what number to include in each cell, one plant (the newest and best automated one) was chosen as a reference plant. Systems managers from FO and PFAT estimated the value in dollar terms of each system in the reference plant. Then a scaling factor was chosen for every other plant to scale up or down the value of the CIM systems in those plants relative to the reference plant. Thus, a plant twice as large as the reference plant might have a scaling factor of two. In each column, for every system installed in that plant, the dollar value for the reference plant was multiplied by the scaling factor to derive the value of each system in the plant. The numbers in each column could be summed up to calculate the value of CIM per plant. These sums in turn could be summed to produce the value of CIM for all of FO.

<table>
<thead>
<tr>
<th>Plant Scaling factor</th>
<th>Reference</th>
<th>#2</th>
<th>#3</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>System A</td>
<td>100,000</td>
<td>150,000</td>
<td>75,000</td>
<td></td>
</tr>
<tr>
<td>System B</td>
<td>75,000</td>
<td>0</td>
<td>56,250</td>
<td></td>
</tr>
<tr>
<td>System C</td>
<td>250,000</td>
<td>375,000</td>
<td>187,500</td>
<td></td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>Total</td>
<td>2,250,000</td>
<td>3,375,000</td>
<td>1,687,500</td>
<td>...</td>
</tr>
</tbody>
</table>
By preparing tables for different years, it was possible to show year-to-year increases in the dollar amount of value contributed to the operation. These results were graphed for presentation at a Systems Efficiency Program meeting.

This technique had the advantage that it quickly produced an estimate of the overall value of CIM to a very large and complex organization populated by many CIM systems. Unfortunately, as the data included in the table were purely guesses made by systems managers located far from the plant floor, the technique also offered no assurance that the resulting numbers were at all meaningful.

The approach taken here can be criticized as another manifestation of Avalon's good news culture. Because the SEP was pushing for hard evidence that its goal of doubling productivity was being met, PFAT responded by preparing a chart like the one shown in Figure 6. The numbers falling out of the five annual tables prepared using this technique conveniently satisfied SEP's goals and showed CIM systems development as being on the right track. There was a danger that had the analysis not produced "good news" results, the analysis and results would have been shelved or reworked to ensure a better outcome.

The data gathering process for the calculations was not clearly and unambiguously defined, objective, hard to manipulate, robust, or repeatable; the data gathering process did not have the characteristics of accuracy included earlier in the definition of good metrics. Thus the metric that was based on this data, the overall value of CIM in dollar terms to FO, was a bad metric. It was one vulnerable to potential misuse in Avalon's culture.
Value Metrics at the Avalon Corp.

The failure of the three approaches of measuring the value of CIM in dollar terms described in this section is a failure that many other approaches developed by many other people in many other firms have also shared. Quantifying CIM's contribution in dollar terms in an accurate, informative, and easy-to-use fashion is an alluring yet unobtainable goal.

The inability to achieve this goal frustrates managers. It makes it difficult to know if past investments in CIM are providing returns in proportion to their risk. It keeps managers from knowing where to invest future CIM dollars most effectively. For CIM projects that included new software development, not knowing how to quantify the value created by various actions during design and implementation makes it hard to guide development towards the most valuable outcome. In spite of this frustration, no general measure of the value of CIM has arisen in wide-spread use beyond flawed financial calculations and customer satisfaction ratings.
Section 4: A Methodology for Creating Value Metrics

While it is often difficult to define good metrics for measuring the value of CIM projects in dollar terms, left open is the question of whether or not there are other good metrics for measuring a CIM team's creation of value. It appears there are. CIM teams can develop a set of project-specific value metrics by using the process-independent methodology defined below. The methodology produces two types of metrics:

1. *Results metrics:* Results metrics can be used after the CIM system has been installed. If the goals set for the metrics are met, then the CIM system will have demonstrably helped to have created value in the plant. The results metrics can also be used by the plant to help continuously improve its operations.

2. *Leading metrics:* Leading metrics can be used during the development of the CIM system to help guide the CIM team towards building systems that create value. For CIM systems that require that new software be written, these value metrics can balance traditional productivity, schedule, cost, and quality software metrics.

Both types of metrics can be used as a part of a successful metrics program. This methodology has been used by a CIM team at the Avalon Corp.; their experience is described below.

**Results Metrics**

Value has been defined as profit, the difference between revenues and expenses. The results metrics produced by the methodology can be tied to profit directly (e.g., by measuring a hard savings like inventory reduction) or indirectly (e.g., by measuring quality improvements that the plant has deemed vital for its future success). A CIM system can be deemed to have helped create value when the goals for the results metrics have been reached (e.g., a 25% inventory reduction or the attainment of a six-
sigma quality level). Note that the CIM system alone does not create all of the value; it is the people and processes in the plant and their use of the CIM system that create value. All should share the credit if the goals of the results metrics are satisfied; they should also share appropriate blame if they are not.

By developing project-specific results metrics, the CIM team should have an easier time demonstrating to the plant and division management that it is creating value. While a financial analysis capturing the direct cost savings or estimates of indirect and opportunity cost savings may also be required (and could be very useful), the goal here is to tie the measurement of value as directly as possible to what the customer has said matters the most.

These results metrics can also be used for ongoing improvement in the plant. Using the examples from above, if inventory levels began to drift up or if quality plunged to three-sigma levels, the results metrics should give warning that corrective action needs to be taken. If plant employees share a philosophy of data-driven continuous improvement, the results metrics should prove useful to them as the plant seeks to maintain the value offered by the CIM system.

**Leading Metrics and the Software Development Process**

Leading metrics can be used to guide the development of CIM systems towards their goal of creating value. Leading metrics generally are only harbingers of value: if the leading metric's goal is satisfied, one can only reasonably expect that when the system is deployed on the factory floor, it will create value.

Leading metrics should be tailored to the CIM team's development process, regardless of whether or not that process includes the production of new software. Some CIM projects require no new software development: for example, some projects deploy purchased shrink-wrapped packages, replicate previously-developed custom-
built software, or configure third-party software packages. Leading metrics for such projects can be useful as management metrics or IT deployment metrics.

But many CIM projects require substantial amounts of new software to be written, either in-house by the CIM team or by a third-party software developer. For such projects, having leading metrics to check the development effort's progress towards having the potential to create value is very useful. The methodology focuses in part on producing such software metrics.

In the past, the lack of leading value metrics created an inability on the part of software engineers to see how their actions would eventually affect the plant's bottom line. This meant that:

- the CIM team could not be sure that it had set the software requirements definition to maximize value,
- during design and implementation, it was difficult for software engineers to know how to make appropriate tradeoffs to maximize value, and
- software development may have been optimized for the other standard software engineering metrics of productivity, schedule, cost, and quality unknowingly at the expense of value. (A metrics mantra at Avalon is "What gets measured is what gets done." If value is not measured, it will not get done.)

Thus, value metrics are vital not only as a measure of IT performance but of software development performance.

**Process-Independent Methodology**

The methodology was designed to be independent of the chosen software development process because metrics alone do not create value; they can only measure and drive progress towards it. The choice and successful execution of a good process for hearing the voice of the customer while specifying requirements and designing the system is what will eventually deliver value to the customer. Many well-defined
processes can help CIM teams accomplish this goal. Unfortunately, many CIM teams today use unstructured, ad hoc processes for listening to the customer and translating that voice into a specification and design. Such processes are not as effective as ones like the following that define structured ways of hearing the voice of the customer:

- **Quality Function Deployment (QFD):** Originally developed in Japan by Toyota, QFD (also known as the House of Quality) was brought to the United States by the Ford Motor Company in the early 1980's. QFD defines a process for matching customer requirements with engineering and manufacturing specifications in order to ensure that the right trade-offs are made so that the most important customer requirements are well satisfied. QFD has proven to be quite effective and its use in new product development has grown rapidly. Recently, QFD has spread to software development; Digital Equipment, Hewlett-Packard, and Avalon report good results from its use there [Hauser and Clausing 1988], [Van Treeck and Thackeray 1993], [Grady 1992], [Hewlett-Packard undated].

- **Joint Application Design (JAD):** JAD was developed at IBM in the late 1970's and its use has been growing rapidly since then. JAD defines a methodology for conducting meetings between developers and users that are supposed to elicit and prioritize a complete list of customer needs. While, as Carmel et al. note, there is little theoretical basis for JAD and not much good, objective research on the topic,

  JAD has become, perhaps, the most common user-involvement methodology in North America for two reasons: First, IS organizations realized that a methodology with a high degree of user involvement would lead to better systems, and they found that solution in JAD; second -- by and large -- it is perceived as working well. The essence of getting the users involved in the JAD methodology is the structured meeting ... The innovation in JAD ... is that the user meeting is structured, disciplined, and is the foundation of the [software development life cycle]. [Carmel et al. 1993]

- **Participatory Design (PD):** While similar to JAD, with its origins in Scandinavia, PD goes beyond JAD in terms of the degree of responsibility and authority given to
users during systems development [Carmel et al. 1993], [Kuhn and Muller 1993]. For CIM teams facing strong, rancorous unions or struggling with quality-of-work issues on the factory floor, PD may be an effective process for bringing the customer into the development of systems that create value.

- **Structured Analysis (SA):** Structured Analysis embodies many different tools designed to make it easier for CIM teams to understand and document a customer's work environment and IT needs. These tools include document flow diagrams (also known as process maps), data flow diagrams, and data structure diagrams (also known as entity-relationship diagrams) [Bull 1989]. Many Computer-Aided Software Engineering (CASE) systems support these SA tools and thus their use is fairly wide spread. While the use of SA does increase the focus on the front end of the software development process and enables CIM teams to translate customer needs and processes into a compact language, Structured Analysis does not specify how the voice of the customer is to be heard or how to ensure that the resulting analyses are accurate reflections of that voice. SA is probably most useful when combined with another process like QFD, JAD, or PD.

While this thesis is about metrics, the thesis asserts that the choice and use of a structured process for hearing the voice of the customer is far more important than the choice and use of value metrics. Replacing an unstructured, ad hoc process with a structured one will do more for the CIM team's ability to create value for its customer than will adding value metrics to either an unstructured, ad hoc process or a structured one.

**Overview of and Motivation for the Methodology**

The goal of the methodology is to rapidly develop a small set of value metrics that can be used to measure and guide the progress of the CIM team towards creating value in the areas that matter most to the customer. The methodology is made up of
nine steps shown in Figure 1 (included at the beginning of the thesis). Each step is facilitated by the use of one or more tools defined by the methodology. The purpose and instructions for each step in the methodology are listed later.

The inspiration for the overall methodology and many of its components was derived from the philosophies of the Total Quality Management movement and from specific TQM tools, notably QFD and Concept Engineering. (See [Shiba et al. 1993], [Guinta and Praizler 1993], and [CQM 1991] for overviews of TQM philosophies, QFD, and Concept Engineering, respectively.) In designing the methodology, the motivation was not only to allow CIM teams to create value metrics but to define a process that would:

- Build into the CIM team an outward focus on the customer.
- Move the CIM team away from "solutions-oriented" thinking and towards "needs-oriented" thinking, per the notions of Concept Engineering.
- Focus the CIM team early on what matters most to the customer, before most development costs are determined.
- Identify non-software obstacles to the creation of value by CIM (such as organizational roadblocks or process problems on the factory floor) that the CIM team and factory must solve before the CIM system can create value. (CIM teams involved in reengineering efforts who have learned the mantra "simplify, then automate" should especially appreciate this.)

The methodology attempts to achieve these higher-level goals by:

- Starting with the customer early in the CIM and software development cycle and returning repeatedly to the customer.
- Using elements of tools like QFD that have well-documented success in turning the voice of the customer into excellent products.\(^\text{10}\)

\(^{10}\) In fact, the first three steps of the methodology could be used to fill in the first column of the first House of Quality.
While there is not yet evidence that use of the methodology will achieve these higher-level goals, it is hoped that as the methodology is used over time, such proof will arise.

**Use of the Methodology**

Before using the methodology, the CIM team must recognize that:

1. As noted earlier, the metrics produced by the methodology will be most powerful when used with structured customer-oriented development processes like QFD. It is the effective execution of these processes that produces value, not the use of value metrics. Therefore, the choice of such a process should be considered a precondition to the use of the methodology (although this is not a formal requirement).

2. The metrics generated by the methodology must be used as part of a metrics program that has the characteristics of successful programs as defined in Section 2. The team must ensure that such a program exists and must construct such a program before running through the methodology.

3. The methodology must be used very early in the software development life cycle when customer requirements are being obtained. The resulting metrics will be most valuable if they are used to guide the team's progress from the outset, not to track its distant past. Early use will also help achieve the higher-level goals described earlier.

Once these conditions are met, the use of the methodology should be led by members of the CIM team. All team members should participate in each step of the methodology for two reasons. Having everyone participate:

1. Emphasizes the importance of focusing on what the customer values.

2. Increases the likelihood that everyone will buy into the chosen metrics. As noted earlier, buy-in is an important part of successful metrics programs.
The CIM team leader should champion and facilitate the use of the methodology. The leader will need to persuade the customer that it is important to allocate time for the interviews, surveys, and presentation required by the methodology. The leader will also need to spearhead continued use of the metrics once the CIM project is underway.

**Definition of the Methodology**

The nine steps in the methodology are described below.
Step 1: Talk to Customers

Purpose:
- To learn what matters the most to the customer in terms of value in the area of the planned CIM project.
- To learn what sorts of improvements the customer wants to make in these areas.
- To start an ongoing dialogue with the customer on the subject of value.
- To listen to the customer -- not to persuade, plan, or negotiate with the customer.

Tool: Structured interviews
- A structured interview is one where a predefined series of questions is asked of the interview subject. By asking multiple subjects the same set of questions in the same order, the interviewer is assured of getting a rich and complete set of responses that can be compared and contrasted.
- The CIM team should prepare a list of open-ended questions designed to spark a torrent of ideas, opinions, and facts from the customer on the subject of what matters most to them in the area of the planned CIM project. Here is a list of sample questions:
  - What changes do you expect the CIM project to create? Why are these things important? Is there a way of measuring these things quantitatively or objectively?
  - What sorts of hard savings do you want? (For example, reductions in headcount, overtime, scrap, or warranty costs; increases in effective capacity; etc.) Can you precisely quantify these goals and the rate at which you want to achieve them?
  - What sorts of soft savings do you want? (For example, increases in customer satisfaction; improvement in product quality; increases in teamwork and proactive
decision-making by front-line workers; etc.) Can you precisely quantify these goals and the rate at which you want to achieve them?

- What do you want from CIM and the CIM team to make these changes happen?
- How would you measure progress towards successfully making the changes you have mentioned?
- Of the changes that you have mentioned, which would create the most value? The least? Why? Which would be the easiest to achieve? The hardest? Why?
- Which of the following aspects of the CIM system are most important to you and why? Functionality, usability, reliability, performance, and supportability.

Note that these are not questions about the design, schedule, features, or feasibility of the CIM project. Instead, they are questions designed to provoke a discussion of the value that is to be created by the CIM project and how this value can be measured.

- Each customer should be interviewed individually by two members of the CIM team. (Interviewing customers as a group runs the risk that the ideas of one or more of the customers will dominate the discussion; speaking to the customers individually avoids this problem although it is more time consuming.11) One CIM team member should act as interviewer while the other should act as a scribe, recording the customer's responses on paper or computer. The interviewer should ask a question, get a full response, and then ask the next question. The interviewer should interrupt the customer only to clarify a point; the customer should be doing almost all of the talking.

- When the interview is done, the scribe should transcribe it immediately on to a computer to ensure that everything the customer said is recorded in full. (The handwritten interview notes may include shorthand or summaries that may be difficult to decode later.) Further, the transcript should include the reaction of the interviewer and scribe to the customer's comments.

---

11 [Griffin and Hauser 1991, p. 9]. This paper includes a quantitative analysis of the trade-offs between individual and group interviews for identifying customer needs.
• The interviewer should encourage the customer to speak factually and specifically. For example, if the customer says that its goal is "to increase machine up-time," the interviewer should attempt to elicit a more detailed response such as "increase machine up-time from today's 70% to 95% as measured by the following formula." This detail will help the CIM team understand exactly what the customer wants.

• The choice of who to include as a customer is important. The team should include as wide a set of customers as is appropriate; a good rule is that if the CIM project will affect:
  - a person's work,
  - the work of some who works closely for them or with them, or
  - if the project needs help from a person to succeed,
then that person is a customer.\textsuperscript{12} Categories of people to consider as potential customers include:

  - \textit{Factory employees:} team members and leaders, line workers, front-line supervisors, area supervisors, shift managers, plant managers, maintenance workers, union representatives
  - \textit{Support employees:} finance, human resources, materials handling, safety, environmental, research and development, marketing, and quality employees
  - \textit{Suppliers}
  - \textit{Downstream customers of the factory:} end customers, distributors, warehousers, value-added resellers

The team should strive for a balance between breadth and depth. To get a complete understanding of what the customer values, people from a wide range of

\textsuperscript{12} Some CIM teams distinguish between customers and users, where customers are "economic buyers" and users are those with hands-on contact with the CIM system. While customers in this context may have a broader view that lets them more easily identify how a system can produce value, the users can still offer insight, especially with regards to obstacles to the creation of value. Thus, we do not use this customers/users distinction in this methodology.
functional areas and positions in the organization's hierarchy should be interviewed. But to make sure that this understanding is accurate, more than one person from each functional area or position in the hierarchy should be interviewed. If the customer population is large, between 20 and 30 one-on-one interviews should identify 90-95% of what matters most to most to the customer in terms of value.¹³

Step 2: Organize and Prioritize Customer Needs

Purpose:

- To structure the ideas gathered in the interviews by grouping them, abstracting them, and then showing the relationships between them.
- To find the customer's most important needs from this structure.

Tool: KJ Diagram

- A KJ Diagram is a technique for grouping, abstracting, structuring, and synthesizing qualitative, fact-oriented information developed by Kawakita Jiro, a noted Japanese anthropologist.\(^{14}\) The steps for developing a KJ Diagram are listed below. The main purpose of a KJ Diagram is to move to a deeper level of understanding of the underlying content and structure of the problem described in the question to be answered by the KJ Diagram. The KJ Diagram allows the team to make sense of a morass of seemingly unconnected facts and lets the team draw conclusions that are well-supported by the facts.
- See Figure 8 for an example of a KJ Diagram.
- The CIM team should create a KJ Diagram to answer the question "What matters most to our customer in terms of value in relation to the CIM project?" The data gathered in the structured interviews should be the basis of the black label statements in the KJ Diagram; this anchors the KJ Diagram with the customer's own voice.
- One great advantage to creating this KJ Diagram is that it can reveal underlying or unarticulated needs that can later by confirmed by the customer.
- Creating a KJ Diagram requires training, practice, and time. In order to become proficient, team members must receive formal instruction on how to create KJ

\(^{14}\) The purpose and use of KJ Diagrams are summarized in [Shiba et al. 1993, pp. 153-169].
Diagrams and then must create several KJ Diagrams in order to hone their skills. But even proficient teams will find that it takes about four hours to create a KJ Diagram.

- Once the KJ Diagram has been completed, the team should post it permanently in a prominent place so that it can serve as an easily accessible reminder of the customer's needs. The team should also transfer the KJ Diagram to the computer so that multiple copies of a more manageably-sized version of the KJ Diagram can be printed and distributed to team members and customers.
How to Create a KJ Diagram

The following list summarizes the steps in creating a KJ Diagram. For a complete description, consult [Kawakita 1991a], [Kawakita 1991b], or [Shiba 1990].

1. **Choose the question** to be answered (for example, "What matters most to our customer in terms of value in relation to the CIM project?"). Explore the environment by talking with customers, gathering numbers, etc. The structured interviews will accomplish this for the CIM team.

2. **Start the KJ Diagram** by having the team write the question in black ink in the upper-left-hand corner of a 2x5' sheet of white paper that has been taped to a wall.

3. **Write fact-oriented information** found while exploring the environment in black ink on yellow Post-It notes. These black labels use the voice of the customer as much as possible.

4. **Scrub the black labels** to ensure that everyone in the team creating the KJ Diagram understands what the labels mean. Rewrite labels to clear up ambiguities or to make the labels more fact-oriented.

5. **Reduce the number** of black labels. Because often there are too many black labels to work comfortably with, the team goes through several rounds of voting to pick the roughly 20 most germane comments. Duplicate comments are often removed.

6. **Group the black labels.** Find common themes between labels and arrange them in groups of no more than three. (Black labels that don't fall into groups are called Lone Wolves and are labeled with a red "LW" to make tracking them easier.)

7. **Write red-label titles** for the black-label groups. These titles should capture the common theme of the group, moving up one level on the ladder of abstraction.

8. **Group the red labels,** just as the black labels were grouped.
9. *Write blue-label titles* for the red-label groups, as before.

10. *Structure the blue labels.* Using causal arrows (⇒) and contradiction signs (←), find causal and contradictory relationships between the blue labels.

11. *Vote* on the most important red labels. This step serves as a point at which the customer is brought back into the KJ Diagram process. Team members and customers review the groups and structure of the KJ Diagram and then cast their vote for the three most important red labels that they see. The top three vote-getters are highlighted on the KJ Diagram to indicate their importance.

12. *Write a conclusion* in black ink in the upper-right-hand corner. Based on the groupings, labels, causal/contradictory relationships, and voting results, the team selects a short concluding sentence to summarize the KJ Diagram.

13. *Sign and date* the KJ Diagram. Write the names of the team members, date, and location where the KJ Diagram was done in the lower-right-hand corner.
Step 3: Verify Results with Customers

Purpose:
- Ensure that the conclusions drawn from the KJ Diagram accurately reflect what the customer values.

Tool: Customer Survey
- The CIM team should send each of the customers interviewed in Step 1 a survey that follows the format listed on the next page.
- The CIM team should average the results as well as plot them on a distribution chart. These numbers will be used in Step 6 to select key metrics.
- If the answers to the overall question indicate that the interviews or KJ Diagram have failed to capture what matters most to all or part of the customers, the CIM team should redo those steps and then repeat this survey. Specifically, if the average response to the overall question was less than 4.5, there is an indication that the team does not yet know what is most valuable to the customer.
- Note that once the KJ Diagram is done, it cannot be altered. While others can disagree with its abstractions, structure, or conclusions, it is not in the spirit of the KJ Diagram to tinker with parts of it once it has been finished. If the customers do not agree with the KJ Diagram, the team should redo it from the beginning.
Customer Value Survey: Sample Format

Project Name & Date

Instructions: Please fill out this one-page survey and return it to nome in the attached envelope by date.

Recently, the CIM team interviewed you and other customers to find out what is most valuable to you with regards to the proposed CIM project. We have summarized what you have told us and now want to make sure that we have heard you correctly and understand what is most important to you. The list below includes what you told us in the interviews plus our groupings and abstractions of what you told us.

For each item on the list, please circle the number that represents how important each item is to you in terms of the value that will be created by the CIM project.

<table>
<thead>
<tr>
<th>IMPORTANCE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not Very</td>
</tr>
<tr>
<td>1</td>
</tr>
</tbody>
</table>

1. Blue label 1
   a. Red label 1
      i. Black label 1
      ii. Black label 2
     ...
   b. Red label 2
      i. Black label 1
     ...
2. Blue label 2
   ...

Overall, how well does this list capture what matters most to you in terms of value in relation to the CIM project?

<table>
<thead>
<tr>
<th>Not Very</th>
<th>Very</th>
</tr>
</thead>
<tbody>
<tr>
<td>Well</td>
<td>Well</td>
</tr>
</tbody>
</table>

If you answered less than 5, why?
Step 4: Think of Metrics

Purpose:
- To develop an exhaustive but unevaluated list of possible metrics for guiding the CIM teams towards delivering systems that create value for the customer.

Tool: Brainstorming
- The purpose of a brainstorming session is to have an uncritical, free-wheeling group discussion to generate ideas -- no matter how outlandish -- on a given topic.
- Here, the brainstorming session should be oriented towards thinking of metrics that will measure how well the CIM team is doing in creating value for the customer, as identified in the structured interviews and KJ Diagram. (This is not a design session.)
- The goal of the brainstorming should be to develop a large number of metrics to evaluate in the next step. The more metrics that are generated here, the less likely the group will have to repeat this step in case an insufficiently broad or appropriate set of metrics is brainstormed.
- To help stimulate thinking along a large number of dimensions, the following questions can be posed to the team during the course of the brainstorming session:
  - Results Metrics: After the CIM system is deployed, how can we measure changes in the areas that matter the most to the customer in terms of value?
  - Leading Metrics: During the early stages of development (gathering customer requirements, defining the functional specification, prototyping), what can we measure to track our progress towards creating a system that will be guaranteed to be useful once it has been deployed?
  - Existing Metrics: How does the customer measure itself already? Can we tap into those measures?
• Developing leading metrics is vital. Lagging results metrics are often easy to identify and to correlate with value but if the team fails to meet its goals for the metric, it is often too late to do anything about it. Therefore the development of leading metrics is very important and the team should consider the following example while brainstorming. Suppose that a priority for the customer is "to get the operators to use the system; in the past they have not because the systems have been too hard to use." This suggests possible results metrics to measure use of the system after deployment: "number of transactions per day," "percent of operators who use the system at least once per shift," or "percent of operators who say that the system is vital to the conduct of their work."

But this also indicates that perhaps the team should focus on ease of use during design. Here, leading metrics would be very useful to the CIM team. If the team decided to build ease-of-use into the system by prototyping the user interface, a good leading metric might be "prototype's ease-of-use ranking by operators on a 1-5 scale." If the operator's ranked the prototype poorly, the team could correct the problem quickly and easily. If they ranked it highly, the team may be able to posit a strong positive correlation between an easy-to-use prototype and eventual use of the working system by factory floor workers.

• Sometimes the best way to create value for the customer is to define new policies. For example, at Chrysler and Hewlett-Packard, customers of the internal CIM organizations complained that it took so long to develop new CIM systems that by the time they were implemented, needs had changed and the systems were obsolete. That is, the CIM systems were not creating any value because they took so long to be developed. In response, Chrysler and HP adopted policies that arbitrarily limited new CIM projects to a maximum of nine and twelve months, respectively. Such a policy could be paired with a metric like "Was the CIM project completed within nine
months? Goal: yes." The CIM team should feel free to brainstorm for policies and related metrics.

- Binary metrics (ones answered "yes" or "no") are acceptable although multi-valued metrics are often preferable. Changes in multi-valued metrics can be tracked over time and can often reveal more than binary metrics. Note that frequently, policy-related metrics will have binary answers for an individual project ("Was the CIM project completed within nine months? yes/no") and multi-valued answers for groups of projects ("What percentage of CIM projects were completed within nine months?").

- Good metrics are often only proxies for value. For example, a CIM project at Chrysler or HP could satisfy the goal of the "Was the project completed within nine months?" metric and still not deliver value to the factory floor. The CIM team could meet this deadline by cutting too many features. But, based on the customer's original complaint, it would be reasonable to assume that delivering projects fast reduces the likelihood that customer needs have changed too much. Thus, satisfying the goals of the metric can be reasonably be assumed to be correlated with delivering value to the customer. Note also that Step 6 of this process ensures that a balanced set of metrics is chosen so that the team cannot improve one metric at the expense of creating value for the customer without hurting another metric.
Step 5: Evaluate the Metrics

Purpose:
- Critique the list of metrics to identify the best ones.
- Generate new metrics by combining or modifying existing metrics to get around problems with the brainstormed metrics that are identified as the metrics are critiqued.

Tool: Evaluation Matrix
- The CIM team should create the Evaluation Matrix as follows:
  - Each of the metrics defined in the brainstorming session should be listed sequentially. The metrics form the rows of the Evaluation Matrix.
  - The columns of the matrix are formed by the following yes-or-no questions:
    1. Is this metric accurate?
    2. Is this metric informative?
    3. Is this metric easy to obtain?
  - The cells in the matrix are filled in by answering each of these questions for each of the proposed metrics. The questions are divided into subquestions, listed on the next page, on which the team should reflect before answering the overall question.
- When all of the metrics have been critiqued, the team should create new metrics by combining or modifying existing metrics with the goal of answering "yes" to all three questions for this metric.
- When done generating new metrics, count the number of "yes" answers for each question and sort the list of metrics in descending order.
Questions to Be Answered in the Evaluation Matrix

1. **Is the metric accurate?**

   Is the metric:
   a. able to accurately measure the target with little random variation or measurement error?
   b. defined clearly and unambiguously?
   c. objective?
   d. hard to manipulate?
   e. successfully field tested?
   f. robust and repeatable (that is, different people measuring the same thing will get the same result)?
   g. sensitive to the right things, insensitive to others?

2. **Is the metric informative?**

   Is the metric:
   a. timely?
   b. easy to interpret?
   c. able to help identify sources of problems and suggest solutions?
   d. able to answer a question about progress towards a goal ([Basili 1992]'s goal-question-metric paradigm)?
   e. based on a model or hypothesis of that which is being measured?
   f. able to be compared with goals, baselines, benchmarks, and control limits?
   g. appropriate to the process maturity of the organization?

3. **Is the metric easy to obtain?**

   Is the metric:
   a. economical? Specifically, does the metric have:
      i. low collecting and reporting overhead?
      ii. automated data collection, whenever possible?
   b. associated with a well-defined data collection process?
Step 6: Select Key Metrics

Purpose:
- Select a small group of key metrics that will help guide the CIM team towards creating value for the customer.
- Set goals, baselines, collection plans, and reporting plans for the chosen metrics.
- Field test the chosen metrics.

Tool: Selection Matrix
- The CIM team should select a small group of balanced metrics that are matched with the customer's needs by creating the Selection Matrix as follows:
  - *Rows and columns*: The rows of the matrix should be made up of the list of needs included in the survey in Step 2. The first column of the matrix should be the average rating obtained for each item on the list in the survey. Subsequent columns should be made up of the top-rated metrics (ones to which each question in the Evaluation Matrix was answered "yes").
  - *Match*: The matrix should be filled in by identifying which metrics are matched with which customer needs and filling in the cell in the matrix with "strong," "medium," or "none." A metric matches a need if by meeting the goals set for the metric, the CIM team will have satisfied that need. Note that more than one metric can match a given need.
  - *Complete coverage*: Once the matrix has been filled in, check to see if all of the needs have one or more associated metrics. If not, develop one or more metrics that match the as-yet unmatched needs by brainstorming and then checking the proposed metric or metrics with the Evaluation Matrix to ensure that they receive all "yes" answers.
- **Cull:** The team should identify the smallest set of metrics needed to match all needs. If the set of matching metrics is large (more than four metrics), attempt to shrink the size of the set by eliminating metrics that match lower-priority needs.

- **Balance:** Once all needs have been matched with metrics, the metrics should be checked for balance. For each need, the team should consider what sorts of bad behaviors the team could exhibit if it wanted to optimize its performance as measured by the metric at the expense of customer value. Then the team should ensure that one or more of the other chosen value metrics or cost, schedule, productivity, or quality measures in use by the team would alert the team to this problem. That is, if the team does well by one metric but is hurting value, another metric should catch the problem.

- Once a small, balanced set of metrics that provides complete coverage of the customer's needs has been defined, the CIM team should set goals, baselines, collection plans, and reporting plans for metrics:

  - **Goals:** Goals should be set such that if the goal is achieved, the matching customer need should be satisfied and thus value should have been created for the customer. If trends in the metric can be measured over time, both intermediate and end goals as well as control limits should be set. The feasibility of achieving these goals can be estimated by examining historical information taken from the CIM organization and customer as well as benchmarking information taken from similar and best-in-class organizations.

  - **Baselines:** Baselines should be set by drawing historical information from the CIM organization and customer.

  - **Corrective action plans:** The team should define how it will react if a metric goes off track.

  - **Collection plans:** A collection plan should define who will collect the data for a metric, how the data will be collected, and when. As much as possible, data
collection should be automated in order to reduce the ongoing costs of metric gathering.

- **Reporting plans:** A reporting plan should define when the metric data will be reported and to whom. Appropriate audiences for the metrics include the CIM team, its management, and the customer. A good framework for presenting the value metrics may be the metrics dashboard discussed in an earlier section. With a dashboard, all of the value metrics are reported together in a one-page graphical format. Complementary metrics for quality, productivity, schedule, and cost can also be included.

- The team should field test the chosen metrics, if possible, to ensure that good data can be gathered as expected.
Step 7: Verify Choices with Customers

Purpose:
- To ensure that the customer agrees that the proposed set of value metrics is a good set.
- To publicly commit the CIM team to using the metrics.
- To continue the dialog with the customer on the subject of value and metrics.

Tool: Presentation
- The CIM team should do the following during a formal presentation to the customer:
  - Summarize the customer's needs.
  - Describe the chosen set of value metrics.
  - Describe how the metrics match with the customer's needs.
  - Summarize the goals, baselines, corrective action plans, collection plans, and reporting plans.
- Comment, criticism, and approval should be invited from the customer. If necessary, the metrics or their goals, baselines, corrective action plans, collection plans, and reporting plans should be changed to address the customer's concerns. If the customer does not approve of the chosen set of metrics because the customer does not believe that they match the customer's needs, it may be necessary to redo this entire process.
- The CIM team should commit itself to the customer to using these metrics in order to help guide itself towards creating maximum value for the customer. The management of the CIM team should do the same.
Step 8: Implement and Use

Purpose:
- Use the metrics to guide the CIM team towards creating maximum value for the customer.

Tool: Metrics and Metrics Program
- As part of a metrics program that has the characteristics of a successful metrics program, the metrics should be gathered and presented per their collection and reporting plans.
- If the CIM team fails to meet one of its goals for a metric, the team should fix the problem by following the corrective action plan appropriately.
Step 9: Improve this Process

Purpose:

- Improve the process of generating and selecting value metrics based on the CIM team's experience in using this process.

Tools: standard TQM problem solving tools (Five Why's, W-V Model, etc.)

- When a CIM project is completed, the CIM team should check that progress towards meeting the metrics' goals was correlated with satisfying the customer's needs. If not, the team should explore why not, using TQM tools like the Five Why's, KJ Diagrams, customer satisfaction surveys, and so on. Based on this understanding, the team should attempt to improve the process of generating and selecting value metrics so that it will work better in the future.

- The CIM team should also look at how this process can be simplified or made clearer and easier to use for future teams. The team should modify this document to pass on that learning.
Advantages of the Methodology

The methodology has several other major advantages in terms of its ability to produce good value metrics. First, regardless of how easily the dollar value of the CIM project can be calculated, the methodology should produce good metrics that are linked to what matters most to the plant in terms of value. Second, regardless of the choice of process for hearing the voice of the customer, the methodology should produce appropriate value metrics. This should be true for any of the structured processes described earlier as well as ad hoc techniques.

The third major advantage is that the methodology should produce a balanced set of good metrics. It draws on the lessons from the literature in an attempt to ensure that only metrics with the characteristics of good metrics as defined in Section 2 are included in the final set. It also attempts to ensure that these metrics are balanced so that they can push the CIM team to maximize value, not maximize a single value metric at the expense of true value.\textsuperscript{15} It attempts to avoid the problems that ad hoc approaches to developing metrics run into that lead to the creation and use of bad metrics.

The fourth major advantage is that the methodology emphasizes the production of leading metrics. The most common value metric, a customer satisfaction rating, is an example of a lagging value metric. It tells the CIM team where it has been -- but not necessarily where it is going. The methodology should encourage the creation of leading metrics that allow the CIM team to take corrective action early before serious mistakes have been made and when it is easier to make changes. These leading metrics

\textsuperscript{15} An interesting example of the importance of balance comes from Xerox. As part of its TQM push in the 1980's, Xerox started measuring its salespeople on the overall satisfaction of each salesperson's customers. The goal was to increase satisfaction. This goal was achieved -- but not making customers happier. Rather, the sales force started to dump its unhappiest customers. While this was logical for the sales force given the incentive structure, it was inimical to Xerox's long-term interests. If the customer satisfaction metric had been balanced with another metric (like "percent of existing customer retained year-to-year"), this perverse behavior would have been limited.
can be tailored to the CIM system's software development process, if new software is being developed, in order to balance the usual productivity, schedule, cost, and quality software engineering metrics. This should help ensure that during the software development process, value is properly considered as the team makes tradeoffs in its attempts to satisfy all of the goals of these sometimes contradictory measures.

**Disadvantages of the Methodology**

The greatest disadvantage of the methodology is that it can produce non-comparable, non-aggregatable metrics. Because the metrics are unique to projects, results from different projects generally cannot be used to make comparisons between projects. Further, results from different projects often cannot be aggregated to give information on how a set of projects is doing or to make comparisons between sets of projects.

For a metrics program driven from the top down, this can be a serious problem. Many firms try to centralize the choice of metrics and push that choice down onto software development teams (for example, this approach was followed at Avalon, Hewlett-Packard, Contel, Motorola, Hughes Aircraft, and Xerox, as detailed in [Avalon 1993e], [Grady and Caswell 1987], [Grady 1992], [Pfleeger 1993], [Daskalantonakis 1992], [Humphrey et al. 1991], and [Xerox 1991]). The metrics reported by development teams are then summed or averaged together to show how the company as a whole is making progress towards certain goals and also to show how various divisions are doing relative to one another. The use of a methodology that produces project-specific value metrics may be inappropriate in such an environment. (Alternatively, however, the project-specific value metrics could be complemented with customer satisfaction ratings that are comparable and aggregatable.)

Another disadvantage is the methodology's bulk. Its size may frighten off some CIM team who might otherwise have been able to benefit from running through it.
A third disadvantage is that the methodology is concerned only with project execution, not project selection. The methodology does not help measure the effectiveness of the IT planning and strategy processes to ensure that out of the universe of potential projects, the ones with the greatest possible value are selected.

The final disadvantage is an agency problem. The methodology gives the CIM team control over the definition and selection of the metrics. While having to get the customer's approval for the chosen metrics in Step 7 helps to alleviate this problem, there still exists a danger that the CIM team will pick metrics that serve its interests and not the customers. (This could be a big issue in environments like Avalon's "good news culture.")

One CIM Project Team's Experience with the Methodology

The methodology was tested on a CIM project being planned for one of the plants in a division of Avalon's Components Operations. This section describes the results of this test and critiques each step of the methodology based on this experience. While the final evaluation of the methodology will have to wait until the project has been completed and while it is clear that improvements to individual stages could be made, overall, the methodology appears to work well. The CIM team and the plant were pleased with the results it produced and believe that the proposed metrics will guide the design and implementation towards a result that will demonstrably create value for the plant.

The proposed CIM system (codenamed "OTD2") was designed to enable the second phase of a large reengineering effort being conducted in the plant. The reengineering program was started in early 1993 in response to a dire crisis: the division was considering halting production of one of the plant's two major products and getting out of that market altogether. Because of gross and continually mounting inefficiencies in the plant's operations, the division was losing substantial sums of
money over this product line. Because the division overall was doing poorly and this product was ancillary to the division's role as an internal Avalon supplier, there was a strong movement at high levels in the division to "rationalize" the business. The plant reacted to this threat by establishing a reengineering program that was made up of four teams. Of these four, only one succeeded.

But it succeeded dramatically. The team was a multifunctional, empowered team charted to reshape the Order-to-Delivery (OTD) process. The team broke its work into two phases, OTD1 and OTD2. In OTD1, over the course of a year, the team changed only plant processes, measures, and attitudes (and made no investments in capital). In doing so, the team succeeded in:

- reducing finished goods inventory for this product line by 90%,
- closing a 244,000 square foot warehouse that had been used to hold all of that excess inventory, and
- cutting order-to-delivery lead time by two-thirds, from weeks to days.

These cost savings and service improvements were very important in the plant's continuing struggle to survive in this fiercely competitive commodity market.

The team moved to Phase 2 in early 1994. The goal of OTD2 is to build on OTD1's successes by improving the scheduling and coordination activities in the plant (from order entry through production to shipping) in order to:

- further reduce lead times,
- lower manufacturing costs, and
- increase the percentage of accurate, on-time deliveries.

The team believes that executing an improved scheduling system is too difficult and too big to do by hand and that, therefore, a new CIM system for scheduling is needed.

John O'Brien, a member of the division's systems department, is heading the design of this OTD2 system. O'Brien has been a member of the OTD reengineering team from its inception and he has contributed his skills in both systems and operations
research. O'Brien's proposal for OTD2 is to develop an artificial intelligence system that will use a set of static and dynamic constraints to produce optimized schedules that will be used to organize activities in the plant. OTD2 is to be developed by a third-party contractor supervised by O'Brien.

Because this will be a large and complex system, OTD2's architecture is designed to allow evolution of the implementation over time. (For an overview of the architecture, see Figure 7.) The plan is to implement the spur line scheduler module, install it, gain experience with it on the factory floor, and then implement and roll out the remaining modules.

![Figure 7: OTD2 Architecture](image)

At the time the methodology was tested, only the general requirements for OTD2 had been defined. There were ongoing discussions between plant and division management over the scope of the project and whether new investment in this product line made sense.

This conflict was one of the system department's motivations for wanting to try out the value metrics methodology. It hoped that use of the methodology would allow it to:
• figure out exactly what the plant needs the most,  
• ensure that the design and implementation satisfy those needs, and  
• gather data to help make the case for approving the project.

Exercising the methodology helped the department achieve these goals.

Details of the Test

The outcome of each step of the methodology are described here. Comment and criticism are included in the next subsection. It took about sixty person-hours spread over three weeks to run through the methodology; another 20 person-hours of customer time were also used. Because the OTD2 project has not yet started full-scale development, the resulting metrics have not been used yet.

Step 1: Talk to Customers

Ten employees of the plant were interviewed by the author for 30-45 minutes each over the course of a single day. They represented a wide range of job functions in the plant and overall business; their job titles were as follows:

area manager  
plant controller  
production control manager  
materials handling supervisor  
production worker (HiLo driver)  
production worker (materials weigher)  
production worker (cutting computer operator)  
capping supervisor  
district sales manager  
sales coordinator

(All of the production workers are union members.) Each of the interview subjects was asked the following questions as part of the structured interview:

1. What are your title and responsibilities?  
2. How long have you worked at this plant? At Avalon?  
3. What is your relationship to the OTD2 project?
4. What changes do you expect the CIM project to create? How will the plant be different? Why are these things important? Can they be measured quantitatively or objectively?

5. What sort of hard savings do you want (e.g., reductions in inventory, headcount, overtime, scrap, customer returns; increases in effective capacity)? Can you precisely quantify these goals and the rate at which you want to achieve them?

6. What sort of other improvements do you want (e.g., increases in customer satisfaction; improvements in quality; increases in teamwork and proactive decision-making by front-line workers)? Can you precisely quantify these goals and the rate at which you want to achieve them?

7. What are the biggest obstacles that the project faces before it can be successful?

Step 2: Organize and Prioritize Customer Needs

A KJ Diagram was prepared by the author based on the data gathered in the interviews; see Figure 8. In answer to the question "What matters the most to the plant in terms of value in relation to the OTD2 project?," the diagram concludes that what matters the most is that OTD2 will give the plant the accurate and timely information it needs to gain consistent control over efficient production, a prerequisite for the plant's survival. The diagram indicates that if the plant could get access to such information and could improve the attitudes of the manufacturing staff, it could spark a chain of events that could dramatically improve the pace and performance of production, make customers happier, cut costs, and eventually aid the plant's survival.

Step 3: Verify Results with Customers

A survey based on the KJ Diagram was distributed the day after the interviews were conducted. The survey is included as Table 8. All of the people interviewed returned the survey in time for inclusion in the summary tabulation except for the area manager and the cutting computer operator. The plant manager and a controls engineer
also filled out the survey. The responses were averaged and sorted; the ranked results are shown in Table 9.

**Step 4: Think of Metrics**

Three hours of brainstorming by O'Brien and the author produced more than 50 metrics suitable for evaluation in the next step.

**Step 5: Evaluate the Metrics**

The brainstormed metrics were evaluated by O'Brien and the author in about two hours to see if they had the characteristics of good metrics. The ordered results are shown in Table 10.
What matters the most to the plant in terms of value in relation to the OTD2 project?

Accurate and timely information will be used to run the business right.

- Accurate and timely production information will be available where needed.
- OTD2 will improve spur scheduling by accounting for all constraints (e.g., sizes, avail workers, # spur lines, etc.)
- Yields will improve by increasing accuracy and timeliness of info flow between packers, truckers, & cutters.
- Spur scheduling system will be flexible enough to handle problems better than the sys that was tried & thrown out 19 years ago.

Measurements will be improved (e.g., in past, grading shifts on yield alone led to unfulfilled orders & long lead times; starting to measure % sched. completed helped to fix).

We will build a "can do," change-oriented attitude in the plant.

- As with OTD1, workers will undergo paradigm shifts (e.g., before OTD1, X-Y% schedules completed was normal. After, Z% is avg! Thinking changed.)
- Without OTD2, people will begin to slip back & to cut corners, esp. with regards to complete orders.
- OTD2 should instill discipline by requiring discipline (don't give an easy out to handle problems).

We will develop consistent control over our business.

- We will not surprise the customer or ourselves.
- Production will be more disciplined and orderly.
- Orders will move predictably and reliably through production.

- Failures to deliver will not surprise the customer, unlike today.
- Improved discipline and preplanning mean OTD2's schedule should be changed only infrequently.
- Sales can give more detailed & accurate order status to customers.

- Accuracy of piece counts in boxes will be improved (tend to undercount today).
- OTD2 should instill greater discipline in managing production flow, esp. capping area.
- OTD2 will give us greater knowledge and control over entire mfg process (when we dump in a new batch, we'll know what truck it'll ship on).

Figure 8 (a): KJ Diagram
Production will be smooth and fast-running.

- Finished goods will flow out of the factory with a minimum of delay and movement.
- Finished goods handling will be reduced, decreasing breakage and saving manpower.
- OTD2 will help reduce order-to-delivery lead time from X days to Y days (best in class level).
- Crates meant for immediate shipment won't be put into inventory first.

Fast, flexible production will help us compete better.

- Cutting lead times is an advantage in a commodity business.
- OTD2 would help reduce the costs of flexible production.
- OTD2 would make records keeping by production workers easier, making them more efficient and happier.

OTD2 will give us the accurate and timely information we need to gain consistent control over efficient production, a prerequisite for the plant's survival.

Happy customers are good customers.

- OTD2 will lead to sustained increases in sales.
  - We can get more customers willing to pay 10-15% price premium for faster, on-time, and accurate service.
  - Cutting and packing bottleneck will be loosened, allowing more gross and net tonnage production & sales.

- Our customers will be happy and loyal.
  - OTD2 would let us gain stable, long-term customer relationships.
  - If we can always hit home runs for our customers (quality products, lowest price, best service), why would they ever go anywhere else?
  - Customer satisfaction (as measured by repeat business and surveys) will improve.

OTD2 would help ensure very accurate and on-time delivery to the ~X% of customers with JIT factories (e.g., framers).

OTD2 will demonstrably cut costs.

- Current X% layout loss will be substantially reduced.
- Fully prepped crates will be delivered just-in-time, reducing crate & dunnage inventory ($X -> $Y; X day supply -> Y day supply) and prep costs.
- Spur lines will be scheduled to reduce manpower needs (X fewer per shift would be great).

Bill Crandall
6 April 1994

It will help the plant survive.

Figure 8 (b): KJ Diagram
Table 8: Survey Questions

1 = Not very important
3 = Somewhat important
5 = Very important

1. Accurate and timely information will be used to run the business right.
   a. Accurate and timely production information will be available where needed.
      i. OTD2 will improve spur scheduling by accounting for all constraints (e.g., sizes, available workers, # of spur lines, etc.)
      ii. Yields will improve by increasing accuracy and timeliness of information flow between packers, truckers, and cutters.
      iii. Spur scheduling will be flexible enough to handle problems better than the system that was tried and thrown out 19 years ago.
   b. Measurements will be improved (e.g., in past, grading shifts on yield alone led to unfilled orders and long lead times; starting to measure % schedules completed helped to fix that).

2. We will build a "can do," change-oriented attitude in the plant.
   i. As with OTD1, workers will undergo paradigm shifts (e.g., before OTD1, X-Y% schedules completed was normal. After, Z% is average! Thinking has changed).
   ii. Without OTD2, people will begin to slip back and to cut corners, especially with regards to complete orders.
   iii. OTD2 should instill discipline by requiring discipline (don't give an easy out to handle problems).

3. We will develop consistent control over our business.
   a. We will not surprise the customer or ourselves.
      i. Failures to deliver will not surprise the customer, unlike today.
      ii. Accuracy of piece counts in boxes will be improved (tend to undercount today).
      iii. Consistency of delivery will be improved (consistency = delivered fast, on-time, and right).
   b. Production will be more disciplined and orderly.
i. Improved discipline and preplanning mean OTD2's schedule should be changed only infrequently. 1 2 3 4 5
ii. OTD2 should instill greater discipline in managing production flow, especially in capping areas. 1 2 3 4 5
c. Orders will move predictably and reliably through production. 1 2 3 4 5
   i. Sales can give more detailed and accurate order status to customers. 1 2 3 4 5
   ii. OTD2 will give us greater knowledge and control over the entire manufacturing process (when we dump in a new batch, we'll know what truck it'll ship on). 1 2 3 4 5

4. Production will be smooth and fast-running. 1 2 3 4 5
a. Finished goods will flow out of the factory with a minimum of delay and movement. 1 2 3 4 5
   i. Finished goods handling will be reduced, decreasing breakage and saving manpower. 1 2 3 4 5
   ii. OTD2 will help reduce order-to-delivery lead time from 7 days to 2-3 days (best in class level). 1 2 3 4 5
   iii. Crates meant for immediate shipment won't be put into inventory first. 1 2 3 4 5
b. Fast, flexible production will help us compete better. 1 2 3 4 5
   i. Cutting lead times is an advantage in a commodity business. 1 2 3 4 5
   ii. OTD2 would help reduce the costs of flexible production. 1 2 3 4 5
   iii. OTD2 would make records keeping by production workers easier, making them more efficient and happier. 1 2 3 4 5

5. Happy customers are good customers. 1 2 3 4 5
a. OTD2 will lead to sustained increases in sales. 1 2 3 4 5
   i. We can get more customers willing to pay 10-15% price premium for faster, on-time, and accurate service. 1 2 3 4 5
   ii. Cutting and packing bottleneck will be loosened, allowing more gross and net tonnage production and sales. 1 2 3 4 5
b. Our customers will be happy and loyal. 1 2 3 4 5
   i. OTD2 would let us gain stable, long-term customer relationships. 1 2 3 4 5
   ii. If we can always hit home runs for our customers (quality products, lowest price, best service), why would they ever go anywhere else?
iii. Customer satisfaction (as measured by repeat business and surveys) will improve.

c. OTD2 would help ensure very accurate and on-time delivery to the approximately X% of customers with just-in-time factories (e.g., framers).

6. OTD2 will demonstrably cut costs.
   i. Current X% layout loss will be substantially reduced.
   ii. Fully prepped crates will be delivered just-in-time, reducing crate and dunnage inventory (X - Y, X day supply -> Y day supply) and prep costs.
   iii. Spur lines will be scheduled to reduce manpower needs (X fewer per shift would be great).

7. It will help the plant survive.

Overall, how well does this list capture what matters most to you in terms of value in relation to the OTD2 project?

Not very well   Very well
1    2    3    4    5

If you answered less than 5, why?
### Table 9: Survey Results

<table>
<thead>
<tr>
<th>Question Number</th>
<th>Average Answer</th>
<th>Survey Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5</td>
<td>Accurate and timely information will be used to run the business right.</td>
</tr>
<tr>
<td>1.a</td>
<td>5</td>
<td>Accurate and timely production information will be available where needed.</td>
</tr>
<tr>
<td>7</td>
<td>4.89</td>
<td>It will help the plant survive.</td>
</tr>
<tr>
<td>2.i</td>
<td>4.89</td>
<td>As with OTD1, workers will undergo paradigm shifts (e.g., before OTD1, X-Y% schedules completed was normal. After, Y% is average! Thinking has changed).</td>
</tr>
<tr>
<td>4.b.i</td>
<td>4.89</td>
<td>Cutting lead times is an advantage in a commodity business.</td>
</tr>
<tr>
<td>6</td>
<td>4.78</td>
<td>OTD2 will demonstrably cut costs.</td>
</tr>
<tr>
<td>1.b</td>
<td>4.78</td>
<td>Measurements will be improved (e.g., in past, grading shifts on yield alone led to unfilled orders and long lead times; starting to measure % schedules completed helped to fix that).</td>
</tr>
<tr>
<td>3.b</td>
<td>4.78</td>
<td>Production will be more disciplined and orderly.</td>
</tr>
<tr>
<td>3.c</td>
<td>4.78</td>
<td>Orders will move predictably and reliably through production.</td>
</tr>
<tr>
<td>5.a</td>
<td>4.78</td>
<td>OTD2 will lead to sustained increases in sales.</td>
</tr>
<tr>
<td>Overall</td>
<td>4.75</td>
<td>Overall, how well does this list capture what matters most to you in terms of value in relation to the OTD2 project?</td>
</tr>
<tr>
<td>3</td>
<td>4.67</td>
<td>We will develop consistent control over our business.</td>
</tr>
<tr>
<td>4</td>
<td>4.67</td>
<td>Production will be smooth and fast-running.</td>
</tr>
<tr>
<td>5</td>
<td>4.67</td>
<td>Happy customers are good customers.</td>
</tr>
<tr>
<td>1.a.i</td>
<td>4.67</td>
<td>OTD2 will improve spur scheduling by accounting for all constraints (e.g., sizes, available workers, # of spur lines, etc.)</td>
</tr>
<tr>
<td>1.a.iii</td>
<td>4.67</td>
<td>Spur scheduling will be flexible enough to handle problems better than the system that was tried and thrown out 19 years ago.</td>
</tr>
<tr>
<td>2.iii</td>
<td>4.67</td>
<td>OTD2 should instill discipline by requiring discipline (don't give an easy out to handle problems).</td>
</tr>
<tr>
<td>3.c.ii</td>
<td>4.67</td>
<td>OTD2 will give us greater knowledge and control over the entire manufacturing process (when we dump in a new batch, we'll know what truck it'll ship on).</td>
</tr>
<tr>
<td>5.b</td>
<td>4.67</td>
<td>Our customers will be happy and loyal.</td>
</tr>
<tr>
<td>6.i</td>
<td>4.67</td>
<td>Current X% layout loss will be substantially reduced.</td>
</tr>
<tr>
<td>6.iii</td>
<td>4.63</td>
<td>Spur lines will be scheduled to reduce manpower needs (X fewer per shift would be great).</td>
</tr>
<tr>
<td>2.ii</td>
<td>4.56</td>
<td>Without OTD2, people will begin to slip back and to cut corners, especially with regards to complete orders.</td>
</tr>
<tr>
<td>3.a.ii</td>
<td>4.56</td>
<td>Accuracy of piece counts in boxes will be improved (tend to undercount today).</td>
</tr>
<tr>
<td>3.a.iii</td>
<td>4.56</td>
<td>Consistency of delivery will be improved (consistency = delivered fast, on-time, and right).</td>
</tr>
<tr>
<td>Question Number</td>
<td>Average Answer</td>
<td>Survey Question</td>
</tr>
<tr>
<td>-----------------</td>
<td>----------------</td>
<td>----------------</td>
</tr>
<tr>
<td>3.b.ii</td>
<td>4.56</td>
<td>OTD2 should instill greater discipline in managing production flow, especially in capping areas.</td>
</tr>
<tr>
<td>4.a</td>
<td>4.56</td>
<td>Finished goods will flow out of the factory with a minimum of delay and movement.</td>
</tr>
<tr>
<td>4.b</td>
<td>4.56</td>
<td>Fast, flexible production will help us compete better.</td>
</tr>
<tr>
<td>4.b.ii</td>
<td>4.56</td>
<td>OTD2 would help reduce the costs of flexible production.</td>
</tr>
<tr>
<td>5.a.ii</td>
<td>4.56</td>
<td>Cutting and packing bottleneck will be loosened, allowing more gross and net tonnage production and sales.</td>
</tr>
<tr>
<td>5.b.i</td>
<td>4.56</td>
<td>OTD2 would let us gain stable, long-term customer relationships.</td>
</tr>
<tr>
<td>5.b.iii</td>
<td>4.56</td>
<td>Customer satisfaction (as measured by repeat business and surveys) will improve.</td>
</tr>
<tr>
<td>5.c</td>
<td>4.56</td>
<td>OTD2 would help ensure very accurate and on-time delivery to the approximately X% of customers with just-in-time factories (e.g., framers).</td>
</tr>
<tr>
<td>2</td>
<td>4.44</td>
<td>We will build a &quot;can do,&quot; change-oriented attitude in the plant.</td>
</tr>
<tr>
<td>3.a</td>
<td>4.44</td>
<td>We will not surprise the customer or ourselves.</td>
</tr>
<tr>
<td>3.a.i</td>
<td>4.44</td>
<td>Failures to deliver will not surprise the customer, unlike today.</td>
</tr>
<tr>
<td>3.c.i</td>
<td>4.44</td>
<td>Sales can give more detailed and accurate order status to customers.</td>
</tr>
<tr>
<td>4.a.ii</td>
<td>4.44</td>
<td>OTD2 will help reduce order-to-delivery lead time from 7 days to 2-3 days (best in class level).</td>
</tr>
<tr>
<td>6.ii</td>
<td>4.44</td>
<td>Fully prepped crates will be delivered just-in-time, reducing crate and dunnage inventory ($X \rightarrow $Y, X day supply $\rightarrow$ Y day supply) and prep costs.</td>
</tr>
<tr>
<td>4.a.iii</td>
<td>4.38</td>
<td>Crates meant for immediate shipment won't be put into inventory first.</td>
</tr>
<tr>
<td>1.a.ii</td>
<td>4.33</td>
<td>Yields will improve by increasing accuracy and timeliness of information flow between packers, truckers, and cutters.</td>
</tr>
<tr>
<td>4.a.i</td>
<td>4.33</td>
<td>Finished goods handling will be reduced, decreasing breakage and saving manpower.</td>
</tr>
<tr>
<td>5.b.ii</td>
<td>4.33</td>
<td>If we can always hit home runs for our customers (quality products, lowest price, best service), why would they ever go anywhere else?</td>
</tr>
<tr>
<td>3.b.i</td>
<td>4.22</td>
<td>Improved discipline and preplanning mean OTD2's schedule should be changed only infrequently.</td>
</tr>
<tr>
<td>4.b.iii</td>
<td>4</td>
<td>OTD2 would make records keeping by production workers easier, making them more efficient and happier.</td>
</tr>
<tr>
<td>5.a.i</td>
<td>3.89</td>
<td>We can get more customers willing to pay 10-15% price premium for faster, on-time, and accurate service.</td>
</tr>
<tr>
<td>Brainstormed metrics</td>
<td>Accurate</td>
<td>Informative</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>1</td>
<td>Want to track indicators of customer behavior that affect your schedule (e.g., # of times the due date and ship date are changed in the order entry system, # crates canceled, # crates rush ordered, # crates canceled close to ship date)</td>
<td>✓</td>
</tr>
<tr>
<td>2</td>
<td>Survey users to ask them if the system is credible and if they would accept its results</td>
<td>✓</td>
</tr>
<tr>
<td>3</td>
<td>Measure if we are scheduling higher yields (# pieces &amp; # crates / # worker hours), not just getting it</td>
<td>✓</td>
</tr>
<tr>
<td>4</td>
<td>Measure # of conflicts between spur lines and layouts (this should be an on-going measure; high # of conflicts indicates a problem in the real world)</td>
<td>✓</td>
</tr>
<tr>
<td>5</td>
<td># of pieces completed / # of worker hours (an overall productivity measure)</td>
<td>✓</td>
</tr>
<tr>
<td>6</td>
<td>Measure user's ability to add new constraints easily and test if they improve things (measure time to add and user satisfaction score)</td>
<td>✓</td>
</tr>
<tr>
<td>7</td>
<td>Fewer people on the spur lines (total # of workers, average #, maximum # could all be measured)</td>
<td>✓</td>
</tr>
<tr>
<td>8</td>
<td>System can accept test data to show that it increases yield and manpower efficiency vs. actual yield and vs. potential yield (using best current practices) on the floor today</td>
<td>✓</td>
</tr>
<tr>
<td>9</td>
<td>Develop coding standard to ensure quality (e.g., no functions with &gt;7 levels of conditionals; limits to # of variables in a function; limits on # of recursive functions)</td>
<td>✓</td>
</tr>
<tr>
<td>10</td>
<td># of pieces completed on-time</td>
<td>✓</td>
</tr>
<tr>
<td>11</td>
<td># of pieces not packed fully the first time through</td>
<td>✓</td>
</tr>
<tr>
<td>12</td>
<td>Lead time (time from order arrives to crates shipped)</td>
<td>✓</td>
</tr>
<tr>
<td>13</td>
<td>% adherence to cut &amp; pack schedule</td>
<td>✓</td>
</tr>
<tr>
<td>14</td>
<td># of orders processed out of sequence</td>
<td>✓</td>
</tr>
<tr>
<td>15</td>
<td>Inventory levels</td>
<td>✓</td>
</tr>
<tr>
<td>16</td>
<td>&quot;Can do&quot; attitude can be measured by a 100% schedule completion rate</td>
<td>✓</td>
</tr>
<tr>
<td>17</td>
<td>% of order status information given to customers that is correct (order life cycle)</td>
<td>✓</td>
</tr>
<tr>
<td>18</td>
<td># of orders sitting on the floor that have been ordered by customer but not received (the Gator report)</td>
<td>✓</td>
</tr>
<tr>
<td>19</td>
<td># square feet of glass in WIP</td>
<td>✓</td>
</tr>
<tr>
<td>20</td>
<td>How long does a box stay in the plant? (% of crates shipped directly vs. going into finished goods inventory)</td>
<td>✓</td>
</tr>
<tr>
<td>21</td>
<td>How long does a box stay in the plant? (% of finished goods inventory that turns over in a day or less)</td>
<td>✓</td>
</tr>
<tr>
<td>Brainstormed metrics</td>
<td>Accurate</td>
<td>Informative</td>
</tr>
<tr>
<td>-------------------------------------------------------------------------------------</td>
<td>----------</td>
<td>-------------</td>
</tr>
<tr>
<td>22 % of trucks that are scheduled when a crate is scheduled</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>23 Employee satisfaction as measured by a survey</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>24 End-customer satisfaction surveys</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>25 Degree to which our order demand cycle moves closer to industry's cycle</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>26 # of customer prototype reviews, signoffs, and changes</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>27 # of errors found during design, coding, and implementation reviews</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>28 # of design, coding, and implementation reviews actually held vs. scheduled</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>29 Measure marginal utility (in terms of increases in yield and improvements in manpower utilization) of adding additional constraints</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>30 Use Extended Relational Analysis; find # of relationships that exceed, say, 3</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>31 OTD2 should enable automation; measure % activities automated</td>
<td>?</td>
<td>✓</td>
</tr>
<tr>
<td>32 Are we putting things into the right spot? (track # of times crates are handled)</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>33 Are we putting things into the right spot? (# of mislocated crates)</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>34 Total sales</td>
<td>✓</td>
<td>×</td>
</tr>
<tr>
<td>35 # hours of unscheduled downtime should drop</td>
<td>✓</td>
<td>×</td>
</tr>
<tr>
<td>36 % of trucks scheduled more than X days in advance</td>
<td>✓</td>
<td>×</td>
</tr>
<tr>
<td>37 Employee absenteeism</td>
<td>✓</td>
<td>×</td>
</tr>
<tr>
<td>38 Have users manually schedule a line and compare their results with automated runs and then show them why the system is better</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>39 Include a counter on how much the system has saved you (at margin and cumulatively) versus the old manual way</td>
<td>×</td>
<td>✓</td>
</tr>
<tr>
<td>40 To measure flexibility, track thickness change time and ribbon-width change time</td>
<td>?</td>
<td>✓</td>
</tr>
<tr>
<td>41 Measure constraints' understandability by asking users and by testing them</td>
<td>?</td>
<td>✓</td>
</tr>
<tr>
<td>42 Accuracy of spur line preventative maintenances should increase</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>43 Package count accuracy rate</td>
<td>×</td>
<td>✓</td>
</tr>
<tr>
<td>44 % of orders delivered right (on-time and accurate)</td>
<td>×</td>
<td>✓</td>
</tr>
<tr>
<td>45 How long does a box stay in the plant? (% of finished goods inventory that turns over in a day or less)</td>
<td>×</td>
<td>✓</td>
</tr>
<tr>
<td>46 Layout loss should be reduced immediately (theoretical vs. actual)</td>
<td>×</td>
<td>✓</td>
</tr>
</tbody>
</table>
Step 6: Select Key Metrics

The top-rated metrics were compared with the customer needs included in the survey by O'Brien and the author. This took about three hours and the resulting matrix is shown as Figure 9. From this matrix, over the course of two hours, a set of key metrics to be used to measure the OTD2 project's progress and results were drawn. The metrics are as follows:

Results Metrics
- Order lead time.
- Finished goods inventory level.
- Number of orders processed out of sequence.

Leading Metrics
- Number of conflicts in the OTD2 system between the spur line and layout schedulers.
- Time taken for and satisfaction of users with adding new constraints to OTD2 and testing if they improve yield and manpower utilization.
• Percentage of scheduled prototype review sessions actually held with customers and number of changes identified therein.

The metrics are described in complete detail in Table 11. The relationships of those metrics alone to the customer needs are included in Figure 10.

Because the selection matrix was so large, it was difficult to see which metrics were tied to the largest number of customer needs. Therefore, O'Brien and the author added a scheme for weighing and ranking the metrics. The weight of metric \( i \) is defined as follows:

\[
\text{weight}_i = \sum_{j=1}^{n} \text{score}_j \times \text{relationship}_{i,j}
\]

where the score is from the survey results and a strong relationship has a value of 4, weak has a value of 1, and none, 0. The resulting weights were then rank ordered to help in picking the set of metrics. This information is shown at the bottom of the Selection Matrix in Figure 9.

These metrics (and their associated goals, baselines, corrective action plans, collection plans, and reporting plans) are shown in Table 11. Note that one of the metrics, order lead time, is intended to be an amalgamation of six of the brainstormed metrics (numbers 12, 17, 18, 20, 21, and 22). The most important overall measure is lead time -- the time from when an order arrives until it ships. But if the average lead time begins to creep up, the lead time needs to be decomposed into its component parts so that the plant will know where to focus its problem-solving energy. Thus, OTD2 can be designed to track the entire order life cycle so that the time spent by orders in production, inventory, and transit can be known. These time measurements of the order life cycle can be used to figure out the overall order lead time as well as to answer questions like "How long does a box stay in the plant?" -- one of the other brainstormed metrics.
Step 7: Verify Choices with Customers

The metrics were presented to representatives from the plant's OTD reengineering team and the third-party software developer in a 45-minute meeting. The author and O'Brien described why value metrics are important, what matters the most to the plant with regards to OTD2, and what the proposed metrics for the project are.

All of the metrics were enthusiastically accepted. However, two of the people in the meeting suggested that the results metrics would not identify productivity problems quickly enough and that another metric like "net tons per person" or "overall equipment effectiveness" (a standard Avalon measure of machine productivity derived from the Total Productive Maintenance movement) should be added to round out the results metrics.

Step 8: Implement and Use

Because full-scale development of the OTD2 project has not yet begun at the time of this writing, there has not yet been the opportunity to use the metrics.

Step 9: Improve this Process

The comments and criticisms listed below are the first attempt to improve the methodology.
## Figure 9: Selection Matrix

| Survey | Score | Metric | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 |
| 1.a    | 5     |        | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● |
| 1.b    | 4.89  |        | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● |
| 1.c    | 4.89  |        | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● |
| 1.d    | 4.89  |        | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● |
| 1.e    | 4.89  |        | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● |
| 1.f    | 4.89  |        | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● |
| 1.g    | 4.89  |        | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● |
| 1.h    | 4.89  |        | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● |
| 1.i    | 4.89  |        | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● |
| 1.j    | 4.89  |        | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● |
| 1.k    | 4.89  |        | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● |
| 1.l    | 4.89  |        | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● |
| 1.m    | 4.89  |        | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● |
| 1.n    | 4.89  |        | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● |
| 1.o    | 4.89  |        | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● |
| 1.p    | 4.89  |        | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● |
| 1.q    | 4.89  |        | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● |
| 1.r    | 4.89  |        | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● |
| 1.s    | 4.89  |        | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● |
| 1.t    | 4.89  |        | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● |
| 1.u    | 4.89  |        | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● |
| 1.v    | 4.89  |        | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● |
| 1.w    | 4.89  |        | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● |
| 1.x    | 4.89  |        | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● |
| 1.y    | 4.89  |        | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● |
| 1.z    | 4.89  |        | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● |

### Legend:

- **●** = strong relationship
- **○** = weak relationship
- **□** = no relationship

### Weight

<table>
<thead>
<tr>
<th>13</th>
<th>15</th>
<th>17</th>
<th>19</th>
<th>21</th>
<th>23</th>
<th>25</th>
<th>27</th>
<th>29</th>
<th>31</th>
</tr>
</thead>
<tbody>
<tr>
<td>13</td>
<td>15</td>
<td>17</td>
<td>19</td>
<td>21</td>
<td>23</td>
<td>25</td>
<td>27</td>
<td>29</td>
<td>31</td>
</tr>
</tbody>
</table>

### Ranking

<table>
<thead>
<tr>
<th>13</th>
<th>15</th>
<th>17</th>
<th>19</th>
<th>21</th>
<th>23</th>
<th>25</th>
<th>27</th>
<th>29</th>
<th>31</th>
</tr>
</thead>
<tbody>
<tr>
<td>13</td>
<td>15</td>
<td>17</td>
<td>19</td>
<td>21</td>
<td>23</td>
<td>25</td>
<td>27</td>
<td>29</td>
<td>31</td>
</tr>
</tbody>
</table>

110
Figure 10: Selection Matrix, Chosen Metrics Only

<table>
<thead>
<tr>
<th>Survey</th>
<th>Score</th>
<th>4</th>
<th>6</th>
<th>14</th>
<th>15</th>
<th>17</th>
<th>26</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>7</td>
<td>4.89</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>2.i</td>
<td>4.89</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>○</td>
<td>●</td>
</tr>
<tr>
<td>4.b.i</td>
<td>4.89</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>6</td>
<td>4.78</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>1.b</td>
<td>4.78</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>3.b</td>
<td>4.78</td>
<td>●</td>
<td>●</td>
<td>○</td>
<td>●</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>3.c</td>
<td>4.78</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>6.a</td>
<td>4.78</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>3</td>
<td>4.67</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>4</td>
<td>4.67</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>5</td>
<td>4.67</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>1.a.i</td>
<td>4.67</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>1.a.iii</td>
<td>4.67</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>2.iii</td>
<td>4.67</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>3.c.ii</td>
<td>4.67</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>5.b</td>
<td>4.67</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>6.i</td>
<td>4.67</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>6.iii</td>
<td>4.63</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>2.ii</td>
<td>4.56</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>3.a.ii</td>
<td>4.56</td>
<td>○</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>3.a.iii</td>
<td>4.56</td>
<td>○</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>3.b.ii</td>
<td>4.56</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>○</td>
<td>●</td>
</tr>
<tr>
<td>4.a</td>
<td>4.56</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>○</td>
<td>●</td>
</tr>
<tr>
<td>4.b</td>
<td>4.56</td>
<td>○</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>4.b.ii</td>
<td>4.56</td>
<td>○</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>5.a.ii</td>
<td>4.56</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>5.b.i</td>
<td>4.56</td>
<td>○</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>5.b.iii</td>
<td>4.56</td>
<td>○</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>5.c</td>
<td>4.56</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>2</td>
<td>4.44</td>
<td>○</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>3.a</td>
<td>4.44</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>3.a.i</td>
<td>4.44</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>3.c.i</td>
<td>4.44</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>4.a.ii</td>
<td>4.44</td>
<td>○</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>6.ii</td>
<td>4.44</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>4.a.iii</td>
<td>4.38</td>
<td>○</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>1.a.ii</td>
<td>4.33</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>4.a.i</td>
<td>4.33</td>
<td>○</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>5.b.ii</td>
<td>4.33</td>
<td>○</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>3.b.i</td>
<td>4.22</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>4.b.iii</td>
<td>4.0</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>5.a.i</td>
<td>3.89</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Weight</th>
<th>7</th>
<th>302</th>
<th>284</th>
<th>340</th>
<th>370</th>
<th>528</th>
<th>355</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ranking</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
</tbody>
</table>
### Table 11: Selected Metrics

<table>
<thead>
<tr>
<th>Results Metric #1:</th>
<th>Order lead time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Goal:</td>
<td>5 days</td>
</tr>
<tr>
<td>Baseline:</td>
<td>8.3 days</td>
</tr>
<tr>
<td>Corrective action</td>
<td></td>
</tr>
<tr>
<td>Plan:</td>
<td>If the metric is trending upwards, conduct a root cause analysis.</td>
</tr>
<tr>
<td>Collection plan:</td>
<td>OTD2 will be designed to track the overall order lead time and all of its components (as defined earlier by the OTD team). Use statistical process control tools like an exponentially-weighted moving average to present the information in a useful manner.</td>
</tr>
<tr>
<td>Reporting plan:</td>
<td>Report this metric monthly to sales, production, and plant management.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Results Metric #2:</th>
<th>Finished goods inventory level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Goal:</td>
<td>Dollar figure to be determined. The sales department currently sets the level of finished goods inventory and needs to be consulted on what the right level of inventory, if any, should be stocked for each type of part.</td>
</tr>
<tr>
<td>Baseline:</td>
<td>Dollar figure to be determined by examining accounting system.</td>
</tr>
<tr>
<td>Corrective action</td>
<td></td>
</tr>
<tr>
<td>plan:</td>
<td>Inventory levels can start to increase for two reasons: (1) orders are filled but not shipped (perhaps because the customer canceled the order) or (2) product is produced to stock and not to order. In either case, sales will have to develop a plan to move the inventory out of the plant.</td>
</tr>
<tr>
<td>Collection plan:</td>
<td>This information is already tracked by the accounting system. This system should be tied to OTD2 so that OTD2 can present that information.</td>
</tr>
<tr>
<td>Reporting plan:</td>
<td>Report this metric daily to the plant manager, area manager, warehouse supervisor, and sales department.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Results Metric #3:</th>
<th>Number of orders processed out of sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Goal:</td>
<td>0</td>
</tr>
<tr>
<td>Baseline:</td>
<td>Not available (there is no real production sequence today).</td>
</tr>
<tr>
<td>Corrective action</td>
<td>If the metric is greater than zero, conduct a root-cause analysis.</td>
</tr>
<tr>
<td>plan:</td>
<td></td>
</tr>
<tr>
<td>Collection plan:</td>
<td>OTD2 will be designed to track this automatically.</td>
</tr>
<tr>
<td>Reporting plan:</td>
<td>This metric should be reported in real time to the plant floor. Also, every thickness change should result in a printed report being sent to production management.</td>
</tr>
<tr>
<td>Leading Metric #1:</td>
<td>Number of conflicts in OTD2 between the spur line and layout schedulers</td>
</tr>
<tr>
<td>-------------------</td>
<td>-------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Goal:</td>
<td>To be determined based on use of OTD2.</td>
</tr>
<tr>
<td>Baseline:</td>
<td>To be determined based on use of OTD2.</td>
</tr>
<tr>
<td>Corrective action plan:</td>
<td>If the metric begins to trend upwards, conduct a root-cause analysis.</td>
</tr>
<tr>
<td>Collection plan:</td>
<td>OTD2 will be designed to track this automatically. OTD2 will have to be integrated with other CIM systems (e.g., the cutting computer) to perform this task.</td>
</tr>
<tr>
<td>Reporting plan:</td>
<td>This metric should be reported when the number of conflicts of their associated dollar cost are high over a period of time. (Use of OTD2 will let us know how high &quot;high&quot; and how long a period we should use.) The conflicts should be broken out by resource (e.g., things under the control of the foreman, engineers, or management) so that they will be notified of things that they can fix individually.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Leading Metric #2:</th>
<th>Time taken for and satisfaction of users with adding new constraints to OTD2 and testing if they improve yield and manpower utilization.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Goal:</td>
<td>To be determined.</td>
</tr>
<tr>
<td>Baseline:</td>
<td>None.</td>
</tr>
<tr>
<td>Corrective action plan:</td>
<td>If it takes too long to add and test new constraints or if users are unhappy with the process, conduct a root-cause analysis.</td>
</tr>
<tr>
<td>Collection plan:</td>
<td>The data should be collected during the prototyping, alpha, and beta stages of the OTD2 software development.</td>
</tr>
<tr>
<td>Reporting plan:</td>
<td>The metric should be reported to the OTD team at standard software life-cycle customer signoff points.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Leading Metric #3:</th>
<th>Percentage of scheduled prototype review sessions actually held with customers and number of changes identified therein.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Goal:</td>
<td>One review session per module (where a module is a screen or a major function on a screen). The expected range of the number of changes will vary depending on the module's complexity.</td>
</tr>
<tr>
<td>Baseline:</td>
<td>None.</td>
</tr>
<tr>
<td>Corrective action plan:</td>
<td>If the percentage of sessions actually held drops below 100%, the plant should consider not signing off on the project. If the number of changes identified during the sessions is above or below the expected number, a root-cause analysis should be conducted.</td>
</tr>
<tr>
<td>Collection plan:</td>
<td>O'Brien will create a form for this metric and track it as project manager.</td>
</tr>
<tr>
<td>Reporting plan:</td>
<td>The metric should be reported to the OTD team at appropriate software life-cycle signoff points.</td>
</tr>
</tbody>
</table>
Critique of the Test and the Methodology

Overall, the OTD2 project proved to be a fair test of the methodology. It revealed the methodology's strengths and weaknesses, laid open some faulty implicit assumptions made during the development of the methodology, and suggested ways of improving the methodology.

The most important of the faulty assumptions was the notion of the CIM "team." The methodology was created with the expectation that it would be used by a small group of systems engineers and managers (say, three to five people total). For many CIM projects at Avalon, having that number of people working at least part time on the early stages of a CIM project is normal. But in the case of OTD2, at this early stage, the team was a single person.

While the author was able to work with O'Brien to form a two-person team for this test, had O'Brien tried to use the methodology on his own, he may have run into some problems:

- Many of the tools (e.g., brainstorming, KJ Diagrams, interviews) tend to be most useful when used by groups. Multiperson use generally brings out a richer and more complete set of ideas. Single-person use can lead to results that better reflect the individual's views than those of the customer.

- One of the methodology's goals was to create a shared vision on the part of the CIM team. If only one person uses the methodology, no shared vision will result. While the artifacts of the methodology (especially the chosen metrics, the KJ Diagram, and the survey results) can be used to teach the people who join the CIM team later about what the customer needs most, instilling a shared passion for providing the customer with value may be difficult without going through a team-based process like this methodology.
While neither of these problems means that the methodology could not be used by an individual, the methodology was not designed with single-person use in mind. Given that many smaller CIM projects are run by a single person in their early phases, this is a large oversight. It suggests that the methodology should be tested by both groups and individuals so that their experiences can be compared. The methodology could then be improved based on that analysis to make it more appropriate for individual use if needed.

The remaining critiques are directed at individual steps in the methodology:

Step 1: Talk to Customers

- While the methodology calls for the interviews to be conducted in pairs by members of the CIM team, the ten interviews were all conducted solo by the author. Because O'Brien had been working hard on the OTD1 effort in the plant for the past year, O'Brien believed that he was too close to the OTD2 project to be objective and that his presence might limit or alter what some of the interview candidates would say. Because candor is important for learning what the customer needs, we decided to have the author go alone and then review the interview results in detail together. These conversations indicated that the interviews were revealing what O'Brien agreed to be the most important aspects of the project's value.

- Getting a broad and deep set of plant employees to interview was vital for uncovering a complete picture of what mattered most to the entire plant. Individuals would often only talk about the OTD2 project's value in relation to their functional area (for example, the sales coordinator emphasized how OTD2 would improve his ability to respond to customer queries about their orders but never talked about how it would improve control over production in the plant; this is a case of "where you stand depends where you sit"). Getting access to an appropriately varied set of customers can be difficult and indicates that the CIM team needs credibility and respect in the plant in order to be able to get people to be
willing to spend time in the interviews. O'Brien's success with the OTD1 effort gave him this respect and he was willing to cash in some of his chips to get people to be interviewed.

- As with all CIM projects in the early stages of planning, there was a degree of fuzziness as to exactly what OTD2 would do. Some thought it was a spur-line scheduler only; others thought it was a complete plant scheduling and order processing system. Members of the OTD2 team were generally aware of the scope of the project; non-members sometimes were not.

  Such fuzziness must be accepted. At this point in a software development effort, an iterative process occurs whereby the customer states its needs and the CIM team states what it can provide. Based on what the team can provide, the customer can modify or reprioritize its needs and the process is repeated. Hopefully, this process will converge on a plan for a CIM system that will satisfy the needs at a good price.

  The purpose of the methodology's interviews is not to resolve this fuzziness but rather to explore what matters most to the customer. These interviews may reveal that what the CIM team had in mind (e.g., implementing the spur-line scheduler first) may not be the best way to maximize customer value. The interview results can be used by the CIM team as they reflect during the iterative definition of the project's requirements.

- Another faulty assumption in this step was that the interviews would elicit a great amount of detail about the customer's needs. The methodology directs the interviewer to "encourage the customer to speak factually and specifically." Often the customer being interviewed did not have hard numbers or specifics (or even good guesses) to put on his needs. For example, in the KJ Diagram, in the group under the label "Accurate and timely information will be used to run the business right," perhaps the most specific label in the customer's own words is the label,
"OTD2 will improve spur scheduling by accounting for all constraints (e.g., sizes, available workers, number of spur lines, etc.)" While these needs are very important to the customer (as confirmed in the survey), they are not particularly detailed. They may not reveal enough information to ensure that the best metrics are being developed or that, if this information is channeled into the software design process, the best design is established.

If the needs elicited in the interviews are too high level ("10,000 foot views," in O'Brien's words), it may be necessary to take the top-rated needs back to the customer and probe for much more detail. Most likely, this should be done as part of the detailed design process anyway. Depending on what the first few trials of this methodology reveal, it may make sense to refine the methodology to include a second level of determining customer needs.

Step 2: Organize and Prioritize Customer Needs

- Preparation of the KJ Diagram went smoothly. Like the interviews, it was created by the author alone because O'Brien felt that he was too close to the project to be objective. Although he had never created a KJ Diagram, O'Brien was critical of the tool because he thought that it was very time consuming and complex and that no one considering using the methodology would be willing to invest the time in creating the diagram.

The author understands these criticisms but believes the value of the KJ Diagram's ability to reveal unexpected insights into complex problems far outweighs the cost of its creation. The author tried to create "simplified" KJ Diagrams with other teams during his internship. This was done by ignoring or modifying the rules for preparing KJ Diagrams with the goal of speeding the process but the results were judged to be less useful than full-blown KJ Diagrams. It appears that strict adherence to the rules (especially the ones limiting the size of
groupings and enforcing the scrubbing of labels), while time consuming, is vital for making the best KJ Diagrams.

Nonetheless, for CIM teams unfamiliar with KJ Diagrams who may be unpersuaded that creating the diagram is worth the trouble, the methodology probably should be changed to include as an alternative, an easier and faster way of organizing and prioritizing customer needs. For example, a technique for writing customer needs on Post-It notes, grouping them in clusters of any size, and abstracting them up two levels would create a three-level hierarchy similar to what is often used as the first column in the first House of Quality.

**Step 3: Verify Results with Customers**

- One unplanned benefit of the survey is that because it takes less than 10 minutes to complete, it can be given to customers who were too busy or were otherwise unavailable for an interview. The OTD2 survey was given to two people whose opinions about the value of the project were very important (the plant manager and a controls engineer) but who were unavailable on the interview day. This ensured that their views were heard.
- The biggest problem with the survey was that the range of responses was too narrow. 92% of the individual responses were four or five (important or very important), as shown in Figure 11. This led to all of the mean responses, save one, being between four and five, as shown in Figure 12. With all of the answers lying between "important" and "very important," it was difficult to distinguish the critical few from the important many.

---

16 The only mean answer less than four was to the statement, "We can get more customers willing to pay 10-15% price premium for faster, on-time, and accurate service." Many of the respondents dismissed the idea that such a premium could be gotten in a commodity market.
This should have been anticipated during the design of the methodology. The KJ Diagram is designed to distill the most important ideas expressed in the interviews so it makes sense that all of the ideas that end up in the diagram and thus included in the survey could be relatively important. Therefore, to cull those critical few from the important many, perhaps the scale should be changed as follows:

1. very unimportant           unimportant
2. unimportant                somewhat important
3. somewhat important ➞ important
4. important                  very important
5. very important             absolutely critical

with no 1-5 numerical scale used lest survey respondents think the responses increase linearly in importance. (Perhaps an exponential scale of 1, 2, 4, 8, 16 would be an appropriate way of weighting the answers.) This scale should spread the distribution of the responses somewhat and should give a better indication of what matters absolutely the most to the customer.

- Another problem with the survey is the phrasing of the last question: "If you answered less than 5 (as an overall rating), why?" This question may bias the respondent towards answering "5" because this answer lets him or her avoid spend time explaining his or her thinking. Thus, the question should be reworked to read "Why did you choose that response?" In this test of the methodology, the answers to this question did not reveal any very interesting information. Perhaps if the new question had been included, more interesting data would have been uncovered.
Step 4: Think of Metrics

- The brainstorming session went well. Going through the KJ Diagram twice, once to think of results metrics and again to think of leading metrics, was useful. No other improvements were identified for this step.
Step 5: Evaluate the Metrics

- It was sometimes necessary to "scrub" the brainstormed metrics to make them clearer, more useful, or more accurate. For example, one of the brainstormed metrics for measuring plant productivity was based on the number of crates produced per day. But in evaluating this metric, it became clear that the number of pieces per crate could vary widely and that, therefore, the piece rate was a better number to include in the metric than the crate rate. The methodology should be changed to make this refinement of the brainstormed metrics more explicit.

- The methodology implicitly assumed that the data necessary for the metric would already be available. This need not be true. If the necessary data is either available today (as assumed) or could be available tomorrow (if the CIM system were designed to gather that information), then the team should be comfortable in concluding that the metric is accurate and easy to obtain. This problem arose several times when the metrics were being evaluated. For example, the data needed for a set of metrics to track customer behavior that affects plant schedules (e.g., the number of times due dates and ship dates are changed in the order entry system) are not available today (other than by a painstaking manual process that has never been tried). But tracking this information could easily be included in the OTD2 system. Therefore, this set of metrics was determined to be accurate and easy to obtain based on the potential availability of the required data. The methodology should make this more explicit.

- Occasionally it was difficult to answer definitively yes or no for one of the attributes. For example, one metric proposed to measure flexibility by tracking thickness change time and ribbon width change time. While there was a consensus that this could be defined clearly and unambiguously and measured robustly, there was disagreement over whether or not the measure was hard to manipulate. Rather than spend the time to resolve such conflicts, the team can mark the attribute with a
question mark rather than a check or an X. It the metric looks like it may be useful or needed in the next step, the team can try to gather more information to resolve the conflict before executing the next step.

- More good metrics were generated than expected, a total of 30. While it is nice to have choices, such a large number of metrics slows down the next step by making the Evaluation Matrix enormous. It might make sense to allow the CIM team to cull the set of good metrics by picking only the best of the good for inclusion in the Evaluation Matrix.

**Step 6: Select Key Metrics**

- Execution of this step produced good leading and results metrics, as desired. The CIM team will be able to use the leading metrics during the prototyping, implementation, and testing phases of the software development life cycle to ensure that it is on track for delivering a system that will be able to help the plant achieve the goals of the results metrics.

  Of the results metrics, only one (finished goods inventory level) is a "hard savings" that is tracked by the plant's cost accounting system. The other two metrics are more measures of responsiveness, flexibility, discipline, and control -- traits that the plant finds very important but that are hard to quantify in dollar terms. But even the "hard savings" inventory metric is a good way to track some of the softer improvements. For example, for:

  - production to be smooth and fast running,
  - finished goods to flow out of the factory with a minimum of delay and movement,
  - finished goods handling to be reduced in order to decrease breakage and save on manpower, and
  - crates meant for immediate shipment not to be put into inventory first,
finished goods inventory will have to be reduced. These improvements will drive not only cost reductions but increases in customer satisfaction, per the KJ Diagram.

- The author and O'Brien selected the leading and results metrics almost as two separate sets of metrics because while they are all complementary, leading metrics are used only before implementation and results metrics, afterwards. The methodology should be changed to make this division and the choice of two sets of metrics clearer.

- The 44x30 selection matrix required that 1,320 cells be filled in. If the list of customer needs could be shortened to include only the very important and absolutely critical needs and if the list of good metrics could also be shortened, it would be easier to complete this matrix. At the same time, this may mean that lesser, though still important, customer needs could fail to be measured or that the absolute best set of metrics might not be chosen.

- As noted earlier, a weighing and rank ordering of the metrics aided the selection process. In the test, the six chosen metrics were all within the top-8-ranked metrics. The weighing formula and ranking process described earlier should be added to this step of the methodology.

- All but six of the customer needs included in the survey have a strong relationship to one or more of the chosen metrics. The ones that did not were:

  7. It will help the plant survive.
  5.b. Our customers will be happy and loyal.
  3.a.ii. Accuracy of piece counts in boxes will be improved (tend to undercount today).
  6.ii. Fully prepped crates will be delivered just-in-time, reducing crate and dunnage inventory ($X \rightarrow Y$, X day supply $\rightarrow$ Y day supply) and prep costs.
  4.b.iii. OTD2 would make records keeping by production workers easier, making them more efficient and happier.
  5.a.i. We can get more customers willing to pay 10-15% price premium for faster, on-time, and accurate service.
The first item on this list, helping the plant survive, was considered to be too abstract to be measured directly by any of the brainstormed metrics as, for the most part, was the second item, making customers happy and loyal. The third item, improving piece count accuracy, was a hot topic in the plant at the time of the survey because a major customer had recently complained about inaccurate counts; fixing that problem is somewhat beyond the envisioned scope of OTD2. The last three items were among the lowest rated needs. Thus, the methodology produced a set of metrics that were clearly linked to nearly all of the most important customer needs.

- Four of the chosen metrics will affect the design of OTD2. Of the results metrics, only finished goods inventory level is tracked automatically by the accounting system today. OTD2 will have to be designed to keep track of the number of orders processed out of sequence and the order lead times and its components.

  For the leading metrics, OTD2 will need to log the number of conflicts between the spur line and layout schedulers. It will also need to provide a mechanism for calculating the improvement in yield and manpower utilization from a set of sample production order data when one or more constraints are added or changed. This will allow the person adding or changing the constraints to know what sort of effect the new constraints will have on these cost drivers.

  These design requirements highlight the value of choosing metrics early. At little cost, the system can be designed to automatically collect and report the data needed to create the measures that will help ensure the system is creating customer value. Adapting the system later to gather these numbers may have been too difficult or expensive to justify the investment.

- Setting goals, baselines, corrective action plans, collection plans, and reporting plans is a critical but time-consuming part of this step. To highlight their importance, these tasks should be broken out into a separate step.
Step 7: Verify Choices with Customers

- One complaint raised during the customer presentation was that the link between the metrics and the customer needs was not clear. The presenters had neglected to include this information in a table in the presentation slides. The methodology should be changed to note that the CIM team should include a matrix that relates the chosen metrics back to the customer needs in the order that they were presented in the survey. This will make it easier for the customers to see how well the metrics are tied to what matters the most to them.

Step 8: Implement and Use

- The metrics have not been used yet so there is nothing to include here.

Step 9: Improve this Process

- These comments indicate at a step-by-step level, there are many improvements that could be made to the methodology. A more holistic view of how the methodology could be improved should wait until the metrics have been used.

Overall Comments on the Methodology

When the methodology was introduced in the thesis, it was stated that the methodology's goal "is to rapidly develop a small set of value metrics that can be used to measure and guide the progress of the CIM team towards creating value in the areas that matter the most to the customer." The test of the methodology shows that this goal can be met. In a matter of days spread over a few weeks, a two-person CIM team was able to assess, organize, and prioritize the needs of a plant considering installing a CIM system for scheduling. The team produced a balanced set of six metrics that could be
used during software development and after installation to ensure that the system satisfies the customer's needs.

Because the metrics have not been used yet, it is too early to say if the metrics will actually be used to affect the software development and plant production processes to ensure that value is created. It is also too early to know if the methodology's higher goals (creating an outward focus on the part of the CIM team, moving towards "needs-oriented" thinking, etc.) will be satisfied.

However, the methodology demonstrates that there are better approaches to assessing the value of CIM than flawed financial analyses or simple customer satisfaction ratings. This should give heart to CIM project managers and manufacturing plant managers that they can develop good objective measures of the contribution of CIM systems to a plant's profitable performance, measures that can be used to guide the behavior of CIM teams and plant employees towards creating value.
Conclusion

Chuck Feltner, whose question, "how do I know I'm doing the right thing?," sparked the research described in this thesis, also proposed an analogy of manufacturing as Gulliver and the support functions (systems, finance, human resources, etc.) as Lilliputians. Each of the Lilliputians, he argued, had thrown a few ropes in the form of bureaucratic restrictions, policies, and practices over the slumbering Gulliver. While none of these ropes individually was strong enough to hold the giant down, together they were. The key to allowing Gulliver to stand, he proposed, was to get each of the Lilliputians to cut the ropes each had used to limit Gulliver's freedom. Once manufacturing as Gulliver had been untied, he would be a fierce force for other manufacturing giants to reckon with.

This thesis has tried to provide the systems Lilliputian with the understanding and tools it needs to break some of the ties that hold manufacturing down. Specifically, the thesis has argued that:

- The persistence in the academic community of the debate over the "productivity paradox" should caution managers that investments in IT do not guarantee good or even adequate returns.
- The inability of these managers to measure the value of CIM in dollar terms in an accurate, informative, and easy-to-use way for most systems is a major obstacle to predicting if a CIM investment will provide an adequate return, to assessing if a past investment did so, and to leading the CIM software development process.
- Beyond customer satisfaction ratings, no other generalized measures of the value of IT have appeared in widespread use.
- When new software is developed for a new CIM system, the lack of value metrics to balance the traditional software engineering metrics for cost, quality,
productivity, and schedule mean that the development process may unwittingly be optimized for one or more of those measures at the expense of customer value.

With this understanding in mind, the thesis provided a new tool, the value metrics methodology, that CIM teams can use to develop metrics tailored to their particular projects.

The methodology produces two types of metrics: results metrics and leading metrics. If the goals of the results metrics are satisfied, it will have been demonstrated that the CIM system, people, and processes in the plant have worked together to create value for the plant in the areas that matter the most to the plant. The results metrics can also be used on an ongoing basis to monitor and improve plant operations.

The leading metrics can be used during the IT development process. For those IT projects that demand the production of new software, these metrics can be used to guide the software development process towards a result that the CIM team is confident could be used effectively by the people and processes in the plant to produce value. These metrics can help the software development team make the right tradeoffs by making value as visible as cost, quality, schedule, and productivity traditionally have been.

The methodology is not perfect, however. It does not produce aggregatable metrics nor does it quantify the CIM system's affect on profit, the definition of value. The system's effect on the bottom line is still an indirect argument. The methodology is bulky and time consuming. It is no substitute for having a good process for hearing the voice of the customer.

But the methodology does demonstrate that it is possible to come up with good metrics that can be used as part of an effective metrics program to measure the value of CIM. These metrics can be used to gauge the success, guide the use, and influence the software development of CIM systems. Simply put, the methodology shows the Lilliputians that there are better ways of knowing if they are "doing the right thing."
Bibliography


