IMPROVING ENTERPRISE DECISION-MAKING: THE BENEFITS OF METRIC COMMONALITY

by

Alissa H. Friedman

Bachelor of Science in Mechanical Engineering
Cornell University, May 2005

Submitted to the Department of Aeronautics and Astronautics
in Partial Fulfillment of the Requirements for the Degree of

Master of Science in Aeronautics and Astronautics
at the
Massachusetts Institute of Technology
June 2010

Signature of Author ____________________
Department of Aeronautics and Astronautics
May 7, 2010

Certified by ____________________
Deborah Nightingale
Professor of the Practice of Aeronautics and Astronautics and Engineering Systems
Thesis Supervisor

Accepted by ____________________
Eytan H. Modiano
Associate Professor of Aeronautics and Astronautics
Chair, Committee on Graduate Students

(this page intentionally left blank)

Improving Enterprise Decision-Making: the Benefits of Metric Commonality

by

Alissa H. Friedman

Submitted to the Department of Aeronautics and Astronautics
on May 7, 2010 in Partial Fulfillment of the
Requirements for the Degree of Master of Science in
Aeronautics and Astronautics

ABSTRACT

The objective of this research is to identify a new approach in managing, and making internal program-level decisions from, externally tracked performance metrics. Industry observations indicate the increasing challenge for program managers and internal development teams to identify performance improvement opportunities for products, services, organizations, etc., in an effective and efficient manner based on tracked performance metrics by external customers. Literature on metrics; performance measurement selection, systems, and frameworks; the concept of commonality; and designing across a life cycle is assessed and helps generate a new concept of commonalizing metrics across an operating life cycle to address this issue. It is hypothesized that despite the uniqueness of each external stakeholder, the tracking of a small set of common performance metrics at different operating life cycle phases across all external stakeholders would result in more accurate decision-making in identifying the most value-added performance improvement opportunities, increased enterprise-level communication, and lower incurred costs.

A detailed case study of a technical product with multiple customers whose external data drives internal program decisions is presented to address (1) if metric commonality is plausible, (2) what the expected benefits are of implementing this new decision-making tool, and (3) how these common metrics would change over the course of the product's operating life cycle. A historical data analysis and initial customer interviews established the architecture of the program's current state. Internal development team expert interviews and a second round of customer interviews were performed in an effort to identify an optimal set of common metrics the external stakeholders could track for this program. Also identified were proper adoption attributes that would need to be considered to not only drive this new decision-making tool through this enterprise, but also to address some of the barriers that influenced the program's current state.

The triangulation of the historical, developer, and customer data sets produced a list of less than a dozen common, value-added metrics for this program, with most of these metrics consistently measured throughout the operating life cycle, supporting the plausibility of this new decision-making tool. Having all stakeholders recording the same metrics also improves the efficiency and effectiveness of making the right product improvement decisions, as well as increases communication within the product community. The study also provides insight into the importance of the voice of the customer, the relationship between metrics and strategic planning, the connection to lean thinking, and a new performance measurement framework; and is considered an excellent starting point for future detailed studies in this area.

Thesis Supervisor: Deborah Nightingale
Title: Professor of the Practice of Aeronautics and Astronautics and Engineering Systems

"If you raise your children to feel that they can accomplish any goal or task they decide upon, you will have succeeded as a parent and you will have given your children the greatest of all blessings."
-- Brian Tracy

Dedicated to my parents, Eileen and Steve.

Another goal accomplished.

ACKNOWLEDGEMENTS

Working on this research over the past two years has provided me with time to reflect on those that have not only helped shape my thesis into the value-added research that I believe it is, but have also helped shape me to be a better researcher, engineer, and person.

To Debbie Nightingale: where do I even begin? To say THANK YOU is quite an understatement! I will always remember the day I was sitting in your office and you told me the incredible news that I would be able to pursue my research interests at my dream school. Fate made you my advisor, but over these past two years, you have become a lifelong friend. Thank you for seeing my potential and for believing in me. I hope I have made you proud.

To Rebecca Boll: many times I refer back to our initial email exchange (on 9/16/08!) that acted as the catalyst for this research. This research would have been but a dream if it weren't for your leadership, support, collaboration, and push to keep me on track. I knew from the beginning that you were a "go-getter" and hope that in my career, I will be as successful as I have seen you be throughout the duration of this study. I sincerely hope I will have the opportunity to work with you again in the future. As such, to John Hassett: thank you for your support in my following through the remaining research when Rebecca assumed her new role.

To Ricardo Valerdi and Jayakanth "JK" Srinivasan: I am consistently wowed by your academic prowess. Despite both of your extremely busy schedules and managing your own students, I greatly appreciate the time you have spent with me to improve my research.

To Dick Lewis: I have learned so much from you, on both a personal and professional level, and hope the two of us can collaborate together on the "next big thing" in enterprise transformation. Thank you for taking an interest in my research and career, and always offering your help; it goes a long way.

To Jorge Oliveira, my "big brother" in the lab: I don't think you even *know* how great of a mentor you are. Despite your extremely busy lifestyle, you take great care in offering advice and guidance to the Masters and rising PhD students in the lab. I have become a better researcher, in part, because of you.

To all the members of LAI: thank you for welcoming me into your home, and letting me be a part of your family. Whether it was collaborating in lab, snowboarding, playing Cranium on game night – thanks for making my time here so fun.

A number of people participated in my interviews. While I cannot name names, please know that my research would not be what it is today without your willingness to participate and your insightful comments. You ARE the research, the motivation, the data – thank you for your cooperation, and I hope my work provides you value.

Similarly, again without naming names, thanks to two separate groups of individuals at work. First, to the group of individuals who have read through, and allowed me to publish, my thesis – thank you for your cooperation and diligence in approving my thesis on time. Second, to my subsection colleagues – I am sure my part-time school schedule over the past three years has not been the easiest thing on the group, but I would not have been able to complete this degree without your support and cooperation. It's things like completing this academic journey that allow me to put into perspective how amazing you all are to work with every day.

But most importantly, thank you to my amazingly supportive family. Thank you for the myriad of phone calls, the cards of support in the mail, and never-fading enthusiasm – the result of which has culminated in a rather large "WOOHOO!" moment, don't you think? Thank you to Jen, Grandma Carolyn, and Grandpa Jack, but more specifically, I would like to thank my parents, Eileen and Steve, without whom I would not have been as ambitious, or successful, in many of my endeavors thus far. I am proud of the woman I am today, and much of who I am has been shaped by your guidance and support. Thank you for believing in me more than I believe in myself (and reminding me each day about potential I don't even know I possess!), and I love you more than I can ever put into words.

…And I'll leave the fighting about who gave me what "qualities" up to you to decide. ☺

(this page intentionally left blank)

TABLE OF CONTENTS

Abstract 3

Acknowledgements 5

Table of Contents 7

List of Figures 9

List of Tables 10

Chapter 1: Introduction 11

- **1.1**: Motivation 11
- **1.2**: Research Question and Hypotheses 15
- **1.3**: Thesis Overview 15

Chapter 2: Metrics – A Literature Overview 18

- **2.1**: The Importance of Metrics 20
- **2.2**: Metric Selection Mistakes 21
  - **2.2.1**: An Unexplored Mistake – Metric Value Changes Over an Operating Life Cycle 25
- **2.3**: Avoiding the Mistakes – Proper Metric Selection 27
- **2.4**: Avoiding the Mistakes – Performance Measurement System Frameworks 30
- **2.5**: A Performance Measurement System Framework for Mature Systems 32
- **2.6**: Applying the Concepts of Product Development Commonality to Metrics 34
- **2.7**: Tying Concepts Together – Case Studies that Begin to Address Metric Commonality 36
- **2.8**: Metrics Literature Conclusions 38

Chapter 3: Research Methods 40

Chapter 4: Proposing the Benefits of Metric Commonality, and its Adoption, through a Case Study 43

- **4.1**: Case Study Background 44
- **4.2**: Initial Research Method and Data Analysis 46
  - **4.2.1**: Metric Data 49
  - **4.2.2**: Customer Email Exchange Data 50
- **4.3**: Identifying Metric Commonality Benefits 53
  - **4.3.1**: Diagnosing Areas of Improvement 53
  - **4.3.2**: Identifying Benefits of Metric Commonality 54
- **4.4**: Recommended Performance Metrics and Adoption Attributes 57
  - **4.4.1**: Performance Metrics Investigation with Developer Experts 57
    - **4.4.1.1**: Expert-Recommended Metrics Analysis 58
    - **4.4.1.2**: Expert-Recommended Metric Commonality Effectiveness and Motivation 60
    - **4.4.1.3**: Expert-Recommended Adoption Attributes 64
  - **4.4.2**: Follow-Up Interviews with Product Customers 66

4.4.2.1: Further Insights into the Current State of the Program 67
4.4.2.2: Comparison of Customer-Generated Metric List to the 69
Historical and Expert-Recommended Metric Lists
4.4.2.3: Customer Adoption Attributes 74
4.5: Case Study Recommendations and Conclusions 76
4.5.1: Product Community Recommendations for Success 76
4.5.2: Case Study Conclusions 79

Chapter 5: Case Study Insights, Reflections, and Impacts 80

5.1: The Importance of the Voice of the Customer 80
5.2: The Link Between Metrics and Strategic Planning................................ 83
5.3: A Framework for Updating a Performance Measurement System with 84
External Input
5.4: Beyond the Case Study – How One Set of Results Can Have a Large Impact . 86
5.4.1: The Connection to Lean Thinking 87
5.4.2: The Importance of Adoption 88
5.4.3: Business Strategy Creation 89

Chapter 6: Conclusions and Future Research 90

6.1: Conclusions 90
6.2: Future Research 92

References 94

Appendix A: Analysis of Strengths and Weaknesses of Performance Measurement 98
Frameworks

Appendix B: Expert Interview Questions 99

Appendix C: Survey on Adoption of Metric Commonality Decision-Making Tool.......101

Appendix D: Customer Interview Questions 102

LIST OF FIGURES

1.1 Simplified Example of Enterprise Architecture 12

2.1 Prasad's (2000) Recommended Measures Across a Product Life Cycle 25

2.2 Sample X-Matrix, Focusing on Metric Alignment 29

4.1 Percentage of International Customers Over the Next Couple of Decades........... 46

4.2 Case Study Stakeholder Value Comparison 48

4.3 Percentage of Common Product Metrics as Presented by the Product Customers ...50

4.4 Research Design to Identify Common Metrics for Case Study 53

4.5 Common Metrics Staying the Same or Varying Throughout a Product's Operating Life Cycle 59

4.6 Expert-Defined Metric Commonality Measurement System Adoption Attributes with 1-Sigma Deviation (n = 8) 65

4.7 Venn Diagram of Shared Metrics Between Historical, Expert, and Customer Data 71

4.8 Expert-Defined and Customer-Defined Metric Commonality Measurement System Adoption Attributes with 1-Sigma Deviation (expert n = 8 and customer n = 4) 74

5.1 Case Study Framework to Create Performance Measurement System 84

5.2 LAI's Four Research Questions 87

LIST OF TABLES

2.1 Metric Elements, as defined by Lohman et al (2004), adapted from Neely (1997).. 30

2.2 Blackburn's (2009) Performance Measurement Framework Typology............... 31

2.3 Key Takeaways from Literature Review Case Studies 38

3.1 Research Method Process, as Defined Using Eisenhardt (1989) Steps 41

4.1 Identifying Benefits of Metric Commonality for Each Area of Improvement 56

4.2 Summary of Historical, Expert, and Customer Data Results 72

4.3 Differences in Expert-Generated and Customer-Generated Adoption Attributes 76

5.1 Common Metric Findings from Case Study 83

CHAPTER 1

INTRODUCTION

1.1: Motivation

Imagine an internal organization, consisting of a program manager and development team, leading the development and deployment of a product based on the wants and needs of a primary customer. This customer has also requested the purchase of "X" products over a given timeframe. This timeframe allows the internal organization the opportunity to continuously improve the product in an effort to increase its operational performance and decrease the assumed maintenance costs by both the customer and the developer. Therefore, a list of performance metrics is created that not only link operating performance to customer strategy and budget allocation, but also provide the internal organization an opportunity to gather data on their product in an effort to continuously improve it. A post-development working relationship has begun.

Now, make the scenario more complex.

While the product was developed for this prime customer, there are other customers in other similar markets that could also benefit from its use. As such, the internal organization markets the product to different customers, and some customers end up purchasing a certain number of the product to be used across a timeframe that works for them. Eventually, the enterprise architecture becomes something as shown in Figure 1.1. The customers using the product are considered stakeholders within the enterprise that includes the internal organization; however, each customer also has stakeholders within its own enterprise with needs that must be fulfilled.

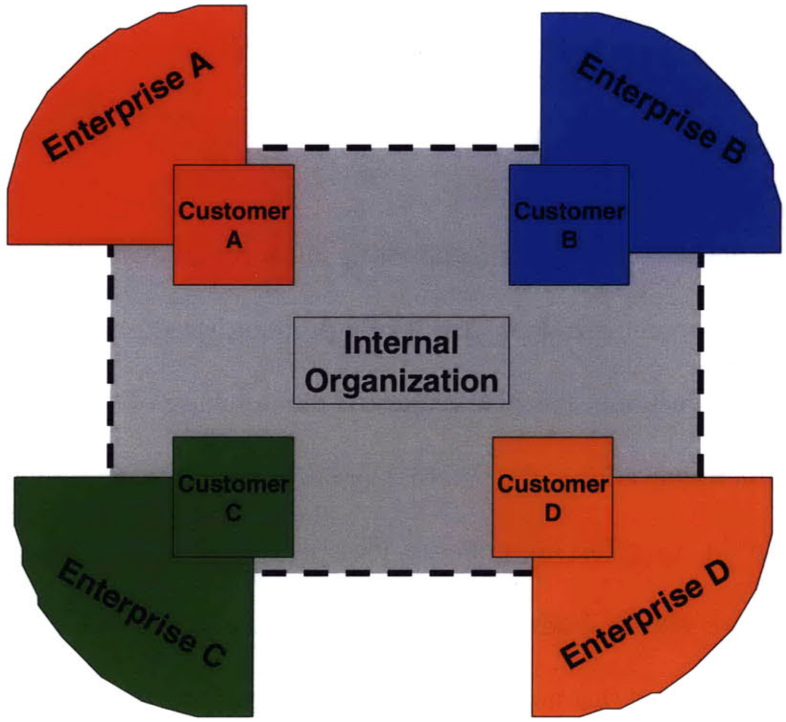

Figure 1.1: Simplified Example of Enterprise Architecture

Currently, multiple customers are operating the product in multiple environments over various timeframes. Despite the common product, the customers are all unique: they do not necessarily value the same things, or make the same decisions, as the prime customer. As such, the original set of performance metrics used by the prime customer is not necessarily the most optimal metrics that should be tracked by these other, newer customers. And while the improvements that are made by the internal organization are based off of the metric data provided by the customers, it is not the internal organization's responsibility to monitor the input, output, or quality of this data. Hence, as time progresses, the program becomes harder to manage because:

- Multiple customers are using the same product multiple ways

- Each customer has their own set of performance metrics
- The product is operating during various life cycle phases, depending on when the product was sold to each individual customer

Therefore, the decisions that the internal organization makes to improve the product may not necessarily be the right decisions for the betterment of the *program* as the metric data is not common, nor is it standardized, across the product community. This results in potential misallocation of budget money for product improvements, a lack of communication between customers, and an increase in the amount of time it takes to digest the data in order to make decisions.

The program manager for this thesis case study puts it best: "Programs should be based on metrics. Much long range planning is based on conclusions and assumptions rather than facts, and not centered on all customer issues."

There must exist a more effective and more efficient way to make the right internal decisions at the right time to improve product performance, and ultimately, customer satisfaction. Understanding that all customers are unique, there must be some level of communication where all customers can maintain their individuality while allowing the internal organization to make the right program decisions to improve their product as a community. Perhaps there is a level of commonality that can be achieved across all customers when tracking product performance metrics.

There are a number of benefits of commonality. Ryan Boas identifies some of these benefits in his doctoral thesis (2008):

> *"Through reuse and sharing, a company may be able to more efficiently develop, produce, and support its products; i.e., produce products with higher lifecycle profits, reduced average lead times, and reduced risks. Total lifecycle costs may potentially be reduced for product families that are based on commonality due to*

potentially reduced development scope, shared economies of scale, and increased degrees of learning, to name a few examples. Revenues may increase when commonality enables a company to produce greater numbers of competitive products."

There is a great deal of commonality in everyday lives, such as the choice of car in which someone drives to work, or the choice of coffeemaker that one chooses to brew a morning cup of coffee. These differences in product makes and models are driven by user needs. Perhaps one of the most famous and successful commonality success stories began a little over a hundred years ago, when Henry Ford introduced the Ford Model T. Each Model T model was built off the same platform. This model of commonalizing features of the car was beneficial to the company, as the car rapidly became the first car owned by many in the country (Alizon et al, 2009). As there were a number of different Model T derivatives, all with common architecture, the platform appealed to a large number of Americans.

Unfortunately, this concept is largely unexplored outside of product development and, as a result, there is growing frustration in making the right program decisions. If common performance metrics are established for a given product across all the product's customers, then areas of product improvement would be more easily targeted, and all customers would result in operating a better-performing product. The same effects would be noted if the right metrics are measured at the right times during the product's operating life cycle. And this example is just for product performance; these same ideas could be applied to assess performance of a service, or even an organization. Despite the uniqueness of one end user compared to another end user, the "thing in question" is common across all users. In order to continuously improve this "thing in question," there should be a common list of metrics (that may or may not vary across an operating life cycle) that could be assessed internally to help drive program decisions. This thesis explores this concept; it addresses the practicality of metric commonality, as well as its applicability across a life cycle.

1.2: Research Questions and Hypotheses

Based on the above motivation, this thesis will attempt to address three questions:

(1) Can the concept of commonality be applied to metrics?

(2) How efficient and effective is commonalizing metrics in assessing performance?

(3) How do common metrics change over an operating life cycle?

It is hypothesized that commonality can be applied to metrics. Perhaps it is harder to apply in the context of this thesis, specifically when the decisions the metrics are used to help drive are dependent on external customer data, but at some level, the concept should be adaptable in this environment. It is also hypothesized that common metrics should more efficiently and effectively help make program-level decisions. Decisions would become more efficient because it should take less time to assess and draw conclusions from common data. Decisions would become more effective because the list of common metrics should be the *right* metrics to track. Finally, metric value should change over an operating life cycle as what is most important during a product's entry into service or during an organization's emerging stage is not necessarily the most important during a product's phase out or an organization's declining stage.

1.3: Thesis Overview

The motivation and problem space for this thesis work has been discussed; the thesis layout is now presented. The layout first begins with theory, briefly outlines research methods used during for the case study, and then shifts to the case study itself. Insights and reflections on the qualitative and quantitative data obtained during the case study research follow, and finally key findings and takeaways are summarized in the conclusion. A more detailed description of each chapter is now provided.

Chapter 2, a literature review on metrics, provides the remaining theoretical detail for this thesis. There is a large body of literature on metrics; specific papers and books that most relate to the motivation for this thesis are explored. This includes the importance of tracking metrics, common mistakes in identifying metrics, and ways to avoid mistakes via micro-level metric selection frameworks and macro-level performance measurement system frameworks. There is also a section highlighting a framework used for creating a performance measurement system when the enterprise already has a system in place that needs updating. This chapter also touches upon commonality literature and its applicability in metric recording, as well as the changing values of metrics across an operating life cycle. It ends with examples that begin to address aspects of this concept, but have their own limitations.

Chapter 3 reviews the research methods used to perform this research. A mixed-method approach was used to support a grounded theory research method in addressing the thesis research questions. Triangulation strengthens the data findings.

The case study is the focus of Chapter 4. The case study identifies a program that shows potential to benefit from metric commonality. The chapter begins with a current state analysis of the program using both historical data as well as customer interview data, based on the case study framework by Yin (2003). It then shifts focus from the current state problem to future state solutions, and identifies the benefits of metric commonality for particular improvement opportunities. A round of interviews with product developer experts and a second round of interviews with the customer contacts results in (1) empirical data highlighting that commonality is achievable in this case study, (2) a discussion on how metrics change throughout an operating life cycle, and (3) the most important attributes the customers believe should be intact to adopt this new decision-making tool. The chapter ends with a list of recommendations for the program.

The case study provides answers to the posed research questions. However, there are other insights and reflections that can be gathered from the study. This gathering of data is the focus of Chapter 5. This chapter speaks to four topics: (1) the importance of the voice of the customer, (2) the link between metrics and strategic planning, (3) the creation of a new framework for updating a performance measurement system that is already in existence, and (4) applicability of results outside the case study context.

Finally, Chapter 6 provides overall conclusions to the thesis, as it reviews the research questions posed in this chapter and answers them, as well as provides areas of further research to continue to address the benefits of metric commonality as a decision-making tool.

CHAPTER 2

METRICS: A LITERATURE REVIEW

"If you can't measure it, you can't manage it." While the origin of this phrase is unknown (web searches list a wide array of potential contributors of the quote), it addresses the importance of metrics for any enterprise. Therefore, it is important to generate the right metrics to assess the performance of something – a product, a service, an organization, etc. – so the right decisions and actions can be made to continuously improve it. However, metric selection is not that easy, and it is only after specifying a particular metric selection process and framework, that an enterprise will be able to understand the connection of metrics and performance.

This chapter reviews literature on metrics that help address this issue. As the chapter progresses, the topics addressed become more specific as a lead-in for the case study in Chapter 4, but still general enough that other enterprises could relate their performance measurement systems with this literature. The chapter addresses the following topics:

(1) Why metrics are important,

(2) Why it is easy to track the wrong metrics,

(3) How enterprises can avoid tracking the wrong metrics,

(4) How enterprises can manage their metrics so they continuously drive the right decisions,

(5) How enterprises can manage their metrics so they continuously drive the right decisions if a performance measurement system has already been established,

(6) How enterprises can manage their metrics so they continuously drive the right decisions if internal decisions are based on external data, and

(7) Examples that attempt to address topics (3) through (6).

While there is a myriad of literature on metrics (in fact, a quick Google search on the word "metric" returns over 33.2 million results), the papers, theses, and books chosen for this chapter specifically address these seven topics.

The chapter begins with a general definition of what a metric is, what its purpose is in an enterprise, and what qualities a good metric should encompass for it to be value-added. Unfortunately, there are a number of common pitfalls that may result in wrong metric selection, which are addressed in the second section. A non-exhaustive list of mistakes is presented to help show how easy it is to measure the wrong thing. The third section begins to address a solution to this problem, as processes are discussed for proper metric selection on more of a micro level. And while metric selection is important, it is also important to understand how to manage all these selected metrics as a system; performance measurement system frameworks are discussed in the fourth section of this chapter to help address this issue (more of a macro-level solution). Performance measurement frameworks are easier to implement from scratch, but likely many enterprises currently have a performance measurement system they would like to improve. The fifth section of this chapter addresses this concern via a case study within the literature.

The overarching limitation in these literature reviews is the absence of a performance measurement system when an enterprise's decisions for, or management of, the improvement of a product or service or organization is significantly dependent on external stakeholder (customer) measures. A new concept of metric commonality is introduced in the sixth section to begin to address this issue. Finally, the chapter closes with two case studies that attempt to address topics (3) through (6), to show both the practical implementations of answers of these questions, as well as the limitations of these case studies themselves.

2.1: The Importance of Metrics

There is a rich body of literature that discusses the importance of metrics and why they are critical for an enterprise's success. Before proceeding through this chapter, it should first be understood what is meant by a metric. Blackburn (2009) highlights differences in metrics from a number of works. He states that some metrics are quantitative in nature, such as cycle time and defect density, while others are more abstract, such as culture and employee involvement. But Blackburn draws one conclusion that is common across all these bodies of literature: a metric quantifies attributes in support of decisions.

There are several qualities that should be used as a guideline for characterizing a "good" metric. Nightingale and Rhodes outline these qualities (2010):

- Metrics are meaningful, quantified measures,
- Metrics must present data or information that allows one to take action;
- Metrics should be tied to strategy and to "core" processes; and
- Metrics should foster process understanding and motivate individual, group, or team action and continual improvement.

All four of these qualities help drive decisions. Decisions are based off of data, which is supported by the first quality of having meaningful, quantified measures. The decision made would support the action identified in the second quality, specifically, what should be done and who should do it. The alignment of metrics with strategy and processes indicates how well the enterprise strategic objectives are being met. Therefore, if there is alignment between metrics and strategies/processes, the enterprise may decide to not make any changes and continue to operate as-is. Conversely, misalignment between metrics and strategies/processes may create a burning platform for change, and a decision would be made that something must be done to realign these three entities. Finally, the fourth quality supports decisions as it acts as a motivator for the stakeholders involved to continuously improve the way they do business.

In order for metrics to be used effectively and help support making the right decisions, a serious effort must be made to properly define them. Unfortunately, this level of effort varies significantly from enterprise to enterprise, and such variation can unfortunately result in a number of metric selection mistakes. While the mistakes addressed below refer more to mistakes on an organizational level, most also apply in the context of this thesis, when internal decisions are based off of external metrics.

2.2: Metric Selection Mistakes

Metric selection mistakes occur all too often, and there is a wealth of literature available that addresses these mistakes. Blackburn (2009) breaks down these reasons into three distinct categories: (1) behavioral effects, or metrics that will influence individual and/or team behavior in an adverse way; (2) value-added, or the metrics that will adversely effect overall enterprise performance and value; and (3) commitment, or the metrics that are not taken seriously enough by leadership. Examples of these mistakes are highlighted below.

Metric Selection Mistakes Relating to Behavioral Effects:

- Measuring metrics that drive the wrong performance (Brown, 1996; Hammer, 2007; Schmenner et al, 1994). For example, employee productivity may be measured by the amount of training classes an employee takes, and as such, employees will therefore spend more time not at their desks taking various training classes that may not be applicable to their job function, knowing that their performance is not affected by whether or not the training actually did help them be more productive.

- Rewarding employees based on behavioral qualities (how many meetings an employee attends, how many stakeholders an employee contacts to solve a problem) rather than valuable accomplishments (Brown, 1996). This is not to say that employee behavior is not important, but perhaps it should not be a measurement criterion to assess enterprise performance.
- Comparing one business unit's performance with another can result in both islands of success and a sense of competition rather than inclusiveness within a company (Brown, 1996).
- Rewarding employees for contributions towards enterprise success in a delayed manner (Hauser and Katz, 1998). The lack of a reward system based on metrics that can measure short-term as well as long-term success may discourage the employee from continuously doing good work.
- Basing employee performance on measures that depend on an uncertain outcome that the employee cannot control (Hauser and Katz, 1998). An employee that contributed beneficial work on a failed program, for example, should not be penalized for having worked on a failed program if the work was still value-added to the company and to the program at the time it was performed.

Metric Selection Mistakes Relating to Metric Value:

- Recording data on too many metrics, as valuable time is wasted as managers (or whomever is assessing the measures) try to sort through lots of data to find the small set of information that is needed for them to do their job (Brown, 1996).
- Choosing metrics that are precisely wrong, but measured because they are easy to measure, or because they address the wrong assumptions (manager preconceptions of what is important to other stakeholders). In essence, managers may believe these measures are related to success, but do not have the evidence supporting this claim. Similarly, measures that make an organization "look good" are used to inaccurately assess the actual performance of the organization, as rather than assessing performance in a truthful manner, performance is assessed using skewed results (Brown, 1996; Hauser and Katz, 1998; Hammer, 2007; Ittner and Larcker, 2003; Schmenner et al, 1994).
- Recording metrics without an appropriate level of detail, making it hard to understand root causes for problems (Brown, 1996 and Hammer, 2007).
- Basing measures on the wrong point of view (Hammer, 2007). For example, an organization may use product availability as a performance measure when they should be measuring how well the product is selling.
- Departments wanting to measure the same thing may use different methodologies, and ultimately different metric definitions, to do so. This results in a loss of validity and reliability when assessing metrics on more of an enterprise-level review (Ittner and Larcker, 2003).

Metric Selection Mistakes Relating to Commitment:

- Enterprises may measure against the wrong strategic objectives. For example, some enterprise data for small and medium-sized organizations focus on measures directly related to short-term success rather than long-term measures that would drive sustainability (Brown, 1996 and Ittner and Larcker, 2003).
- Making decisions on metrics that do not address the intent of recording metrics in the first place, or, "losing sight of the goal (Hauser and Katz, 1998)."
- Organizational boundaries dictate performance metrics (Hammer, 2007).
- Some leadership is "not serious about measurement in the first place. It is manifested by arguing about metrics instead of taking them to heart, by finding excuses for poor performance instead of tracking root causes, by looking for ways to pass the blame to others rather than shouldering the responsibility for improving performance. If the other errors are sins of the intellect, this is a sin of character and corporate culture (Hammer, 2007)."

In addition to these three categories, there is an unexplored area: selecting the right metrics to assess performance *at the right time* in an operating life cycle. Consider the life cycle of a human being. The quality of life of a child is different than his or her adulthood. Similarly, an elderly person values different aspects of life than someone thirty years his or her junior. The consideration of metric value across a life cycle should also be dictated in the metric selection process. The section below discusses this in terms of a product, but the concept of metric value changes across a life cycle can be considered for whatever the "something" being tracked actually is.

2.2.1: An Unexplored Mistake – Metric Value Changes Over an Operating Life Cycle

Metrics should change, both in its quantifiable results as well as value to the enterprise, throughout a life cycle of a product, service, organization, etc. Life cycle design is important to take into consideration.

In terms of a product, the goal of a life cycle design is to maximize the value of a product, while containing its impact on cost to manufacturer, the user and the society to a minimum (Prasad, 2000). These changes in types of measures are highlighted in Prasad's paper, and shown in Figure 2.1.

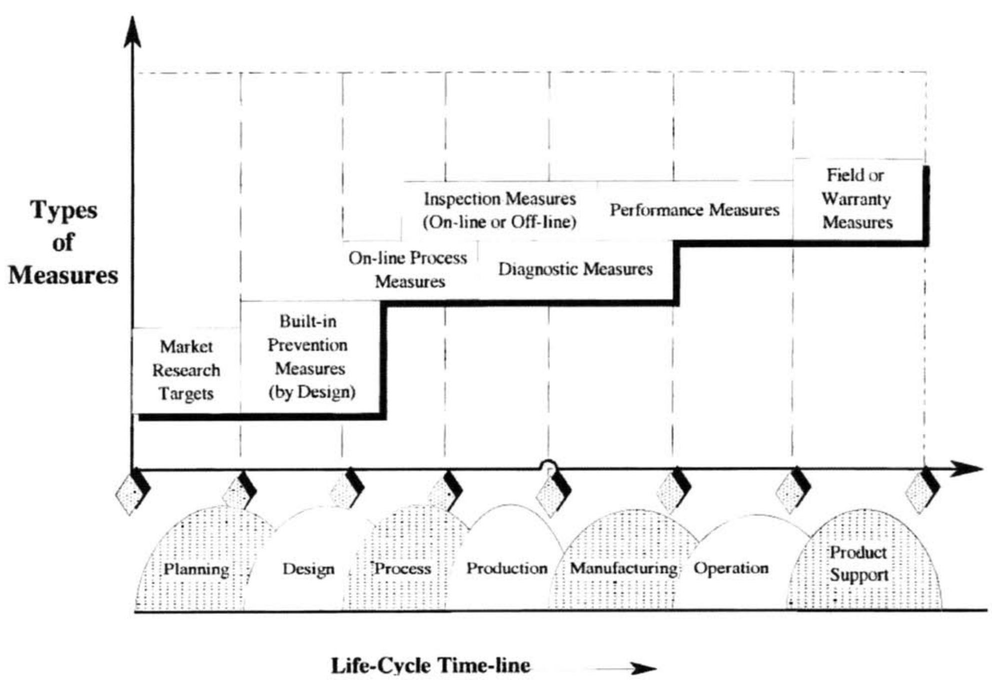

Figure 2.1. Prasad's (2000) Recommended Measures Across a Product Life Cycle

Unfortunately, there is a literature disconnect about generating metrics across the lifecycle; specifically, that operational and product support metrics shown above are not given enough thought during the pre-operative life cycles of the product. Prasad highlights that most engineering and design teams do not go far enough in reducing these life cycle driver costs. Most focus on the company costs and in a narrow sense just concentrate on the direct costs (such as labor and materials), and that few teams attack the company's greatest cost challenge – the indirect costs (Prasad, 2000). Perhaps part of the reason why there is a disconnect between these life cycle phases is because there is no incentive to think about metrics at an enterprise level, as usually products are handed off to the user once they are developed and business progresses, as usual. At the time of a product's entry into service, the user requirements that were previously defined have been met, and the product is deployed. However, there is value in tracking product performance after this deployment in an effort to alleviate maintenance costs and provide the customer with a high quality of the product. This, in turn, also increases the reputation of the product developer if the product developed is continuously improved. If the intention is to lower overall product cost, then more effort should be given to creating and assessing the right metrics during the operation and product support life cycle phases. This includes the cost incurred during the operation and product support phase of the product. A case study paper that analyzed software metrics supports this statement; the authors stated more research in the area of specific metrics for the maintenance phase of the life cycle is necessary, since two-thirds of the cost of a software system is spent on maintenance (Cote et al, 1988).

While this addresses the concept of identifying metric value changes that address a product during its operational life cycle, this is not a boundary. Nightingale (2009) highlights changes in metric value across other kinds of life cycles. A business goes through the organizational phases of emergence, growth, maturation, and declination. As such, metrics that assume more value at these different phases are cash flow, competitive advantage, market share, and critical mass, respectively. Also listed is the entity of core competency – recognition, learning, practice, and expertise. Metrics that change in value for this example are inventory of skills and capabilities and competitive advantage; acquiring knowledge and cycles of learning; use, application, levels of use in an organization, and deployment; and teaching, leveraging advantages, and combining and evaluating; respectively.

The metric mistake in this section is the making of decisions from a metric that does not align with the particular life cycle phase in which it is being assessed. Making these wrong decisions would not maximize the value of a product, service, or organizational performance, nor would it contain impact on cost to other stakeholders, because the wrong thing is being measured at the wrong time.

2.3: Avoiding the Mistakes – Proper Metric Selection

Mistakes can be minimized, or eliminated, if metrics are methodically thought out. Nightingale (2009) identifies four steps that should be followed when selecting metrics:

(1) Relate metrics to value and supporting decisions;

(2) Identify what you know, need to know, and the value of the information;

(3) Determine how metrics impact behavior and align with organizational levels; and

(4) Create systematic processes, feedback, and a measurement-friendly culture.

Nightingale also asserts that the metrics should satisfy three criteria: they should be strategic, quantitative, and qualitative. The strategic aspect aligns the metric, and addresses the question of how to align the metric, with enterprise strategic objectives. The quantitative aspect is the data itself that the enterprise can analyze to determine its current status in meeting these objectives. Finally, the qualitative aspect is to ensure that the metric provides value to the enterprise; that is, that it does not act as a "false alarm" (Schmenner et al, 1994) as its measurement does not provide the right value to the enterprise.

A more visual way of assessing if a metric is considered a "good" metric is by using the X-Matrix tool from the MIT Lean Advancement Initiative toolset (Nightingale et al, 2008). The X-matrix allows the enterprise to assess if a given strategic metric is assessed by a given metric (the strategic aspect), and if a given metric measures a given key enterprise process (the qualitative aspect).

The corresponding cell that aligns the strategic objective and metric, or metric and key process, is highlighted blue (or darkened) for a strong correlation, yellow (or marginally darkened) for a weak correlation, or white for a lack of correlation. This provides a representation of the alignment of these metrics with the rest of the enterprise. Note that this tool supplements the third quality of a "good" metric, as it ties the metrics to the strategies and core processes of the enterprise. Figure 2.2 is an example of the metrics portion of an X-Matrix. This was used during a spring 2009 class presentation on a portfolio acquisition enterprise; data has been normalized.

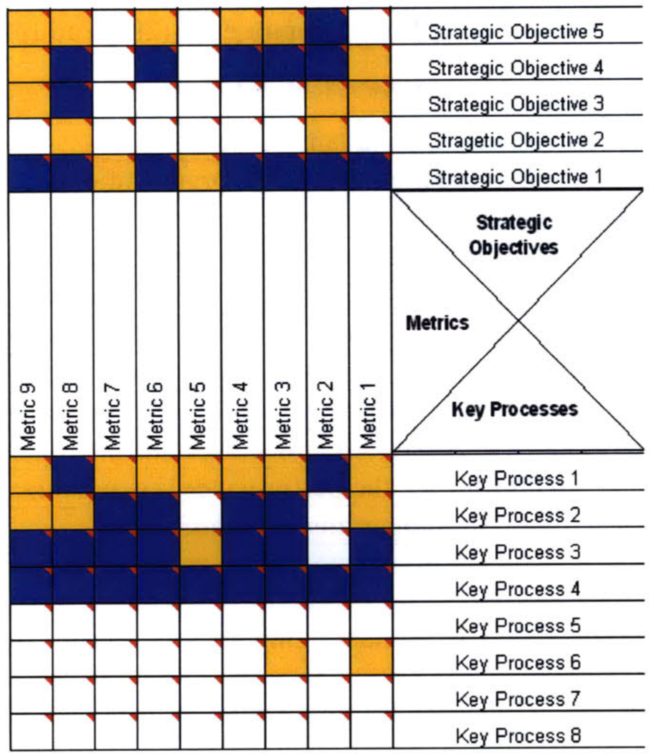

Figure 2.2: Sample X-Matrix, Focusing on Metric Alignment

There are a number of frameworks that can be used to select the right metrics; however, all metrics should have certain elements that help address the previously mentioned characteristics. Lohman et al (2004, adapted from Neely, 1997) defines these elements in Table 2.1:

Table 2.1: Metric Elements, as defined by Lohman et al (2004, adapted from Neely, 1997)

Metric Attribute	Explanation
Name	Use exact names to avoid ambiguity
Objective/Purpose	The relation of the metric with the organizational objectives must be clear
Scope	States the areas of business or parts of the organization that are included
Target	Benchmarks must be determined in order to monitor progress
Equation	The exact calculation of the metric must be known
Units of Measure	What is/are the unit(s) used
Frequency	The frequency of recording and reporting of the metric
Data Source	The exact data sources involved in calculating a metric value
Owner	The responsible person for collecting data and reporting the metric
Drivers	Factors that influence the performance, i.e. organizational units, events, etc.
Comments	Outstanding issues regarding the metric

This is also an excellent first step in standardizing metrics across different stakeholders in different organizations, as this framework (or an adaptation of it) will allow all stakeholders to begin to "speak the same language."

2.4: Avoiding the Mistakes – Performance Measurement System Frameworks

While it is important to use a framework in selecting metrics and standardizing their definitions via a breakdown of their metric elements, it is also important to understand the importance of overall performance measurement. Performance measurement is defined as the process of quantifying the efficiency and effectiveness of an action (Neely et al, 1995), where efficiency is "doing things right," and effectiveness is "doing the right things (Drucker, 1963)." Therefore, performance measurement systems are sets of metrics used to quantify both the efficiency and effectiveness of actions (Neely et al, 1995). More specifically, they are collections of financial and non-financial performance indicators that managers use to evaluate their own or their unit's performance or the performance of their subordinates (Tuomela, 2005). Information derived from performance measurement systems can also be used for resource allocation, coordination, business evaluation, and early warning identification (Tuomela, 2005 which references Simons, 1995). Performance measurement systems therefore serve a number of purposes.

Because there is much diversity in how these systems can help provide "order" to an enterprise, there is no universally correct performance measurement system. Fortunately, there have been a number of different frameworks developed over time, all with different foci. Blackburn (2009) provides a comprehensive list of a dozen different frameworks that are structural (typology-based), procedural (methodology for establishing the system), or a combination of both. His list is provided in Table 2.2.

Table 2.2: Blackburn's (2009) Performance Measurement Framework Typology

Structural	Procedural	Both
Strategic Measurement & Reporting Technique (Cross et al., 1988)	A Framework for Design & Audit (Medori, 2000)	The Balanced Scorecard (Kaplan et al., 1992)
The Performance Prism (Neely et al., 2001)	A Framework for Factors Affecting Evolution (Kennerly et al., 2003)	Extended Enterprise Balanced Scorecard (Structural) and Procedural Frameworks (Folan et al., 2005)
European Foundation for Quality Management – EFQM (Jackson, 2001)	Define-Measure-Analyze-Implement-Control (De Feo et al., 2005)	-
PSM's Measurement Contstruct (McGarry et al., 2001)	GQM (Basili et al., 1994)	-
Value Stream Mapping (Murman et al., 2002)	Steps to Metric Selection	-

Similarly, Mahidhar (2005) provides strengths and weaknesses for half of these performance measurement frameworks (see Appendix A). Within all of these different frameworks, however, there is a common limitation: the connection of external metrics driving internal decisions. That is, if internal decisions are based off of external data, how does an enterprise align its internal system with its external one? This unanswered question will be explored in the case study for this thesis.

2.5: A Performance Measurement System Framework for Mature Systems

Most of these frameworks are rather practical when a performance measurement system is to be created from scratch; the challenge, therefore, lies in taking these frameworks and implementing them into an enterprise with an already established performance measurement system. This is surprising, as "businesses rarely want to design PMS's from scratch. Usually managers are interested in eliminating any weaknesses in their existing system (Neely et al, 1994)." There is also little literature on addressing a performance measurement system outside an operations environment. As noted in Lohman et al's (2004) case study paper: "there is far less literature that provides an understanding of how the process of developing a performance measurement system is impacted by existing measures (or new measures that are being developed simultaneously as a result of other initiatives) at various levels both within and outside the operations function." One paper in particular addresses this issue via a "performance measurement audit" and includes a case study that uses this framework.

Medori and Steeple (2000) lay out a performance measurement framework, which includes both financial and non-financial measures, for an already-established enterprise. The enterprise in his case study is a manufacturing organization, but manufacturing operations are not being assessed. Instead, this case study focuses on organizational performance. Medori breaks down the performance measurement framework into six steps, with one step set up specifically for programs with already-established performance measurement systems. The six-step process is as follows: (1) identify company success factors, (2) use a performance measurement grid to align measurement capability with strategy, (3) select measures using a checklist, (4) perform a measurement audit, (5) implement the measures, and (6) periodically maintain the measurement system.

The first step develops the company's mission and strategy, and associated business objectives that address the overall strategy. The second step aligns the strategy identified in step one with measurement categories. This particular example identifies the categories as quality, cost, flexibility, time, delivery, and future growth. Step three aligns metrics from a pre-populated list of financial and non-financial metric definitions with the metric categories defined in the performance measurement grid in step two. At this point, the company has a list of metrics that they feel are value-added to track.

The fourth step addresses the issue of updating, or revamping, an established performance measurement system of an established enterprise. This is important as current literature reviews highlight that little or no consideration is given for existing measurement systems that companies may have in place (Medori, 1998). The audit capability is a framework that effectively identifies if a company's existing measurement system is current and measuring what is valuable to the enterprise. The audit is broken down into three separate decision criteria. If existing measures are aligned, or are the same as, the new selected measures, they are kept and continuously used. If the existing measures are not aligned with the new selected measures, and do not provide the enterprise with value, they are classified as "false alarms" and should be eliminated. Finally, if there are remaining new measures that do not align with existing measures, these remaining new measures should be implemented. This last set of measures are defined as "gaps," or, measures that are important to a company's success but are presently not being measured by the company's measurement system (Medori and Steeple, 2000 which referenced Dixon et al, 1990). Note that this step could be skipped entirely if the company is already trying to create a performance measurement system from scratch.

The fifth framework step is the actual implementation of these value-added measures. The company uses a list of metric elements to help establish the metric itself, similar in nature to the framework highlighted by Lohman, and Neely, in Table 2.1. In the case of this study, the eight metric elements are the metric title, objective, benchmark, equation, frequency, data source, responsibility, and improvement. The sixth, and final, step in this framework is the periodic maintenance of these measures, a predetermined operating rhythm that the right company stakeholders would reassess metrics and propose updates, if necessary. In essence, the audit step would need to be repeated in the future.

While this is one example of the updating and refinement of an already existing measurement system, it is one procedural framework. What is necessary for all enterprises faced with this issue is some sort of "audit" process that works for them. A framework with similar intentions but different procedures is applied to the thesis case study in Chapter 4; its results, and the questions that were asked to the stakeholders, are then reviewed.

2.6: Applying the Concepts of Product Development Commonality to Metrics

As discussed in Chapter 1, there are enterprises that make decisions based off of external stakeholder, or customer, metric data. The inability to manage this data could result in making wrong decisions, spending more time than necessary in making these decisions, and an overall lack of communication between the customers and other stakeholders. Literature on commonality may help answer this metric problem.

Commonality is most often referenced in product development. It results from reuse of assets that were previously developed to meet the needs of another product and, in some cases, from reuse of assets that were specifically developed to meet the needs of multiple products (Boas, 2008). There is a great deal of commonality with the products that are used everyday. Take the Sony Walkman during the timeframe when a Walkman was similar in commodity as today's iPod as an example of a commoditized item that benefited from product commonality. Sony uses a strategy of a combination of novel technological advances, along with tweaks of existing designs, to target models for distinct market segments (or, understanding customer needs). Sony also relies on flexible, automated manufacturing processes to keep the costs of these changes to a minimum (Sanderson and Uzumeri, 1995). This allows more models to be marketed to more users and provides Sony with a strong share in this market. There are a number of case studies that highlight the benefits of commonality in product development. However, there is little literature, if at all, about using the concept of commonality outside the product development community.

The commonality definition as stated by Boas can therefore be rewritten, just slightly, to have it address external metrics instead of product development. The definition would become: the reuse of metrics that were previously developed to meet the needs of another customer/external stakeholder and, in some cases, from reuse of metrics that were specifically developed to meet the needs of multiple customers/external stakeholders. If all external stakeholders use the same product, have a similar organizational goal, or have product commonality on a tangible level, then there can be some level of common metrics that can be used for all external customers and leveraged by the internal enterprise to make decisions. The case study in Chapter 4 begins to address this concept.

2.7: Tying Concepts Together – Case Studies That Begin to Address Metric Commonality

As mentioned earlier, the concept of metric commonality and commonalizing not just what to measure, but how to measure things, can be applied outside of a product or service. Two examples will be discussed in this section that utilize aspects of this idea.

The implementation of standardized metrics greatly improved the European Operations department at Nike. In this situation, Nike wanted to improve its supply chain enterprise and generated a list of metrics appropriate for their corresponding business objectives, yet the definitions of the metrics varied depending on who was using them. Nike therefore initiated a "metrics dictionary," with approximately 100 metrics and definitions agreed upon by operational managers. Both internal Nike communication and metric quality improved as a result of every user speaking the same language and comparing the same things throughout the operations department (Lohman et al, 2004). Additionally, Nike was able to create a scorecard, with three levels of metrics, which can be assessed by different stakeholders at different times to help drive decisions. Yet, there are limitations within this case study: first, while all metrics were standardized, there were still many metrics that are used to help drive decisions, and perhaps this number could be minimized and still be efficient, and just as effective, in helping drive decisions through the organization. This case study also did not address an already existing performance measurement system; it built one from scratch. Additionally, this was created for an organization with internal stakeholders; it does not outwardly address the concept of external stakeholder metrics to help drive internal stakeholder decisions.

A similar case study, the creation of the Commercial Aviation Safety Team (CAST) and its impact to flight safety, provided a quantifiable result to commonalizing metrics. The CAST team, comprised of industry and government safety experts, formed in 1997 and set a goal to reduce US commercial aviation fatal accidents by 80%. What was necessary was an understanding of what safety data existed that could be used as a leading indicator of fatal accidents. Prior to this team formation, there was no universal standard for safety data, thus making development of a common safety agenda rather difficult. As this team developed common taxonomies and definitions for aviation accident and reporting systems, it became easier to both interpret data and communicate within the aviation community (Stephens et al, 2008). As stated in the Stephens article, the absence of a common taxonomy and the lack of industry data-sharing initiatives greatly diminishes the ability to recognize emerging risks and increasing threats before their manifestation in an accident or serious incident. In a sense, the CAST turned a reactive community into a proactive one by having all stakeholders speak the same safety language. Improvement programs in the form of safety enhancements were introduced to lower the risk of fatal accidents. The 10-year outlook reduced the risk of fatal accidents by 83% (Interview with CAST member, 2009 Oct 8). In this example, both external stakeholders and the metric data they record were used in the creation of a standardized database of safety definitions. However, what was not discussed was the sheer volume of metrics that were created, what metrics were actually used to drive safety decisions, and the importance of various metrics across the life cycle of the products that generated this data. Additionally, this example also created a performance measurement system from scratch and therefore did not "audit" a current system already in place.

Table 2.3 summarizes the key points and takeaways from these two studies.

Table 2.3: Key Takeaways from Literature Review Case Studies

Case Study	Nike European Operations Department	Civil Aviation Safety Team
Reference	Lohman et al, 2004	Stephens et al, 2008; Interview with CAST team member
Objective	Improve supply chain management via integration of various local performance indicators into company-wide performance management system.	Reduce the US commercial aviation fatal accident rate by 80% over 10 years via the use of a standardized taxonomy of safety metric data.
New PMS?	Yes	Yes
Timeframe	6 months	9/1997 – 2005 at least
Stakeholders Involved in Metric Creation Process	Management	Industry and government safety experts from ICAO, air carriers, airframe and engine manufacturers, pilot associations, regulatory authorities, transportation safety boards, members from North America and Europe
Results	• Metrics scoreboard with high, medium, and low level metrics • Metrics dictionary with ~100 standardized definitions across business	• Earlier identification of emerging risks and increasing threats • The transition from a reactive to proactive system
Successes	• A set of standardized metrics	• Standard definitions generated by internal and external stakeholders (sensitive data normalized by 3rd party)
Limitations	• Did not address how the 100 metrics were prioritized for decision-making • Methods were not applied on already-established performance measurement system • External stakeholders were not a part of the performance measurement system development process	• Did not address how the sheer volume of standardized metrics drove safety decisions at various points in time • Methods were not applied on already-established performance measurement system

2.8: Metrics Literature Conclusions

The topics presented in this chapter drew on each other in an attempt to address specific issues that many enterprises face today. Metrics are very important to track, as they help stakeholders make decisions on, and generate actions, to improve performance of "something" – where that "something" can be a product, service, an organization, etc. Unfortunately, there are a number of mistakes that are made all too often when trying to select the right metrics for that "something," and are broken down into mistakes based on behavior, value, and commitment. One mistake that was not addressed in the literature, not selecting the right metric at the right time in an operating life cycle, was also addressed. Fortunately, there are a number of metric selection elements and frameworks that can be followed to help minimize the opportunity to make these mistakes, though there are some limitations in the literature that currently exists on this topic. Specifically, there is a lack of literature on the development of a performance measurement system when a performance measurement system already exists; regardless of the situation, a "performance audit" should be performed to identify (1) what metrics should continue to be tracked, (2) what metrics should be eliminated as they no longer add value, and (3) what metrics should be added to provide more value in making decisions. The Medori and Steeple (2000) case study begins to address this. A more unexplored area in metric creation is the creation of a framework when internal organizational decisions are driven by external metrics, an idea that may be solved by applying the concept of product development commonality into the metrics world. A couple of examples that most closely relate to the concept of metric commonality were discussed, but each example had its own limitations in a broader environment.

The Chapter 4 case study will further explore this area. The study will review most of these concepts within its own boundaries, as well as relate to the concept of metric commonality across an operating life cycle.

CHAPTER 3

RESEARCH METHODS

This research uses a mixed-method approach to address the research questions in Chapter 1. There is a combination of qualitative and quantitative data obtained, and analyzed, throughout the case study. The research performed largely follows the method of grounded theory, which is "an approach which is based on the systematic development and refinement of categories and concepts from the collected data in order to build theory, such that the final outcome is a theory that is 'grounded' in data (Haddad, 2008)." That is, the thesis began with the development of research questions based primarily on observations, and reinforced by gaps in literature. The data obtained through the study helped generate theory that supports the initial observations.

It should be noted that this paper does not include the raw data that was collected throughout the research, so as to protect the enterprise. Additionally, the interviewees are not addressed by name or title. All interview are stripped of any specificity, and any visual depiction of data collection is normalized as well.

There are four rounds of data collection throughout this research and, as the research progresses, the questions the data addresses becomes more specific, also supporting the grounded theory approach. A review of historical presentations within the context of the case study helps generate the current state analysis of the case study. The first round of interview questions that immediately followed that data collection process help to address *why* the case study got to where it is today. This first round of customer interviews are created using Robert Yin's case study methods (2003) and will be discussed further in Chapter 4. The expert interviews that follow are developed to begin to answer the research questions. The second round of customer interviews are developed using insights from the expert response data and historical data.

All expert and customer interviewees were specifically chosen for this research, and all have provided their consent to participate in the study.

The research method follows the process detailed in Eisenhardt (1989). It is outlined in Table 3.1.

Table 3.1: Research Method Process, as Defined Using Eisenhardt (1989) Steps

Step	Thesis Research
Getting Started	Research question reactively defined: based on workplace observations and gaps in literature reviews
Selecting Cases	One program chosen that exhibits problem, and specific (non-random) choosing of interviewees based on expertise and roles
Crafting Instruments and Protocols	• Historical presentations will provide initial quantitative data, and customer interviews will provide initial qualitative data, to generate current state analysis • First round of expert interviews and second round of customer interviews will provide qualitative and quantitative data
Entering the Field	• Create customer interview questions using Yin's case study methods (2003) • Correlate historical data with customer interviews to generate current state analysis • Create expert interview questions on current state analysis • Create 2nd round of customer interview questions on current state analysis and expert responses
Analyzing Data	Triangulation: correlate three sets of mixed data using graphs, charts, and Venn Diagrams. Support quantitative data with qualitative data.
Shaping Hypotheses	Correlate data and its current insights to original research questions. Refine the context of the research question, if necessary.
Enfolding Literature	Eisenhardt (1989): "What is this similar to, what does it contradict, and why?"
Reaching Closure	Anticipated answers to research questions

As mentioned earlier in this chapter, three sets of data, all within the same dimension of a research problem, are compared with each other to help draw theory and insights from the case study. This approach, therefore, utilizes triangulation as a way to strengthen the validity of the results and the mixed-method research design. This case study hopes to address the concept of metric commonality across an operating life cycle, as well as the effectiveness and efficiency gained or lost in doing so. The three sets of data used to support these questions are: (1) historical data from presentations (quantitative data), data obtained through expert interviews (both qualitative and quantitative data), and data obtained through customer interviews (both qualitative and quantitative data). Thus, three different viewpoints are used to address the same concepts in the following chapter.

CHAPTER 4

PROPOSING THE BENEFITS OF METRIC COMMONALITY, AND ITS ADOPTION, THROUGH A CASE STUDY

The focus of this chapter is to present an enterprise that has the potential to benefit from adopting metric commonality as a new decision-making tool, and to determine the attributes that would help this enterprise adopt it. This chapter is organized into five main sections: (1) the enterprise background and motivation for change, (2) the initial case study data analysis that helped to define the enterprise current state, (3) the identification of the metric commonality benefits specific to this case study, (4) the final case study data analysis that piece together overall recommended metrics and adoption attributes, and (5) recommendations and conclusions.

The case study background provides a brief history of the enterprise and begins to shape the enterprise's current state. The motivation for change is created by fast-forwarding a few years, when the enterprise will go through a significant shift in stakeholder value and will need to reconsider the way business decisions are made.

The initial case study data in the third section is generated following Yin's case study structure and methodology. A number of interviews with selected enterprise stakeholders, as well as data analysis through internal stakeholder presentations to the community, both confirm and continue to refine the current state of the enterprise.

The third part of this chapter transitions from identifying the issue at hand to identifying its solutions. A new research design is introduced and followed. The current state analysis serves as the catalyst to identify overall areas of improvement for this case, and the benefits of metric commonalty for each area of improvement are proposed and discussed.

The fourth part of this chapter uses the results of its preceding section as a foundation to identify overall recommended metrics and adoption attributes for this enterprise. Two sets of data are reviewed in this section. A list of expert-recommended common metrics across different phases of a life cycle, as well as recommended strategies for the adoption of a new decision-making tool, are analyzed, based on a number of interviews with experts in the case study field. The results of this interview set are then used as a baseline for a second round of interviews with the same stakeholders of part two of this chapter, obtaining the same kind of data from a different perspective.

The fifth and final section of this chapter completes the case study with a list of recommendations and case study conclusions.

4.1: Case Study Background

This case study examines the performance metrics of a technical product. This enterprise is considered a prime candidate for adopting metric commonality across the product's operating lifecycle. This product, originally developed for a large, US-based primary customer, was introduced into its operating environment decades ago. Since this time, the product has been sold to other customers, all of them international. This product, of which thousands are now currently in operation, is considered a global product. Its resulting enterprise architecture resembles something similar to Figure 1.1. Over the past few decades, as the number of operating products increased, so did the number of performance metrics tracked by the various customers. Presently, as the number of recorded metrics continue to increase, it becomes harder for the product community to understand and manage its operating performance, as well as to recommend suggested improvements, and it also becomes more challenging for the customers to compare the performance of their product against other customers in an effort to identify which customer is best-in-class. This results in much long-range planning for the product program based on assumptions rather than actual performance data. Additionally, the program is skewed towards the needs of the primary customer, and not necessarily the newer, international customers.

In an effort to improve the performance of the product, it is necessary to understand how the product performs in its various operating environments, and what drivers can be improved upon to further enhance its operation. As such, the customers are invited to participate in a periodic conference that focuses on identifying these improvements. Currently, a majority of the customers participate, and present the metrics that they track. While there is some consistency in high-level metrics, not all of these metrics are measured the same way, or use the same metric definition. Additionally, assessing the root causes that would yield trends in the higher-level metrics also proves to be an arduous task, as there is a lack of commonality across all customers in what is measured, as well as *how* things are measured. What could be hypothesized, in this situation, is the process of identifying the proper improvement opportunities may not be as effective or efficient as it could be. Additionally, the presentations of different customer metrics during these conferences results in more time and effort (and thus, more customer- and company-funded money) needed to compare the same things against each other.

This issue is currently a catalyst for change for the product community, and will only become more so over the next few years, as the primary customer will begin to phase out this product. Meanwhile, the other customers will continue to operate the same product in their own operating environments, with an anticipated longevity of at least another two decades. Figure 4.1 highlights the increase in product use by these customers over this timeframe.

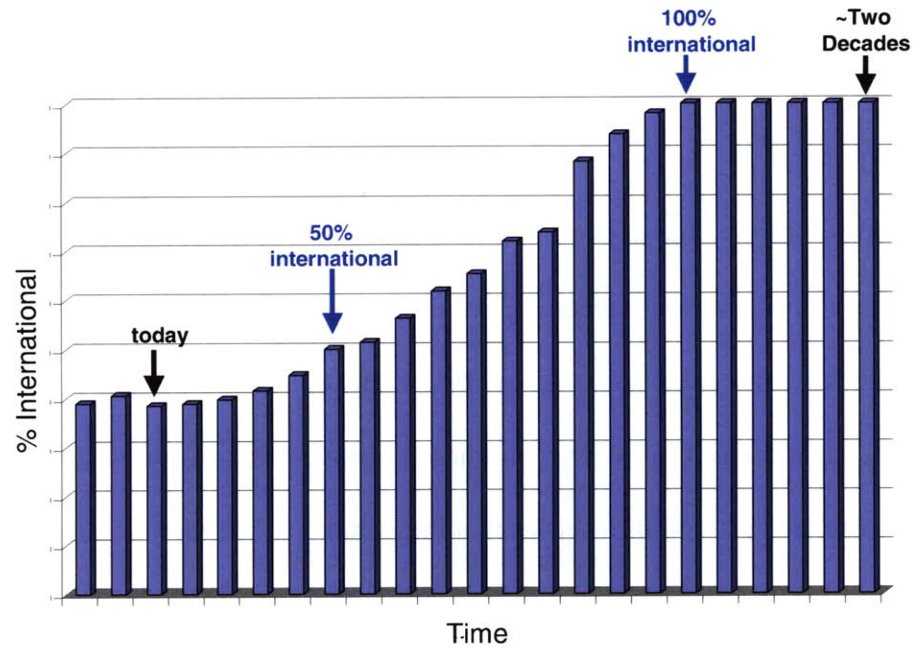

Figure 4.1: Percentage of International Customers Over the Next Couple of Decades

It is necessary to establish a better way of managing product performance. Generating an optimized set of common metrics across all customers would allow the program manager to better assess product performance and focus on the right improvement programs to continually improve its operation, and it would allow customers to be able to communicate with other product users as well as identify which customer is best-in-class. It is first necessary to perform a current state analysis and understand how the program diverged over time.

4.2: Initial Research Method and Data Analysis

This case study began by following Yin's case study structure and methodology. Yin's five components of a research design are:

(1) its questions;

(2) its propositions, if any;

(3) its unit(s) of analysis;

(4) the logic linking the data to the propositions; and

(5) the criteria for interpreting the findings.

The overarching question to answer was: why have these performance metrics diverged over time? There were a few propositions that could answer this question:

(a) Customers have unique metrics that are input into and output from different kinds of databases;

(b) Customers that are more knowledgeable of the product or have more capability to record data will have a more exhaustive list of performance metrics;

(c) A driven program manager who can appropriately lead, manage, plan, influence, support and react will help drive the collection and analysis of proper performance metrics;

(d) Varying levels of technological advances can be associated with the monitoring, recording, and assessment of data; and

(e) Not all customers will want to share their data with other customers.

The unit of analysis largely came from email exchanges with the customers themselves, though quantitative data was also used. As mentioned earlier, most of these customers gather for periodic conferences, where their performance data is shown to the product community. This quantitative data also supports the case study. The logic linking data to the propositions came from the customer email exchanges. The email exchanges and quantitative metric data were used as the criteria to interpret these findings, as inconsistencies in the customer email exchanges as well as significant differences in the customer metric presentations support the propositions.

The product customers served as the main stakeholder for this case study. A stakeholder value comparison (Figure 4.2) for this enterprise shows the customer not only as the most important stakeholder relative to the enterprise, but also as the stakeholder who anticipates the most value from the enterprise (as the customer is the end user of this product).

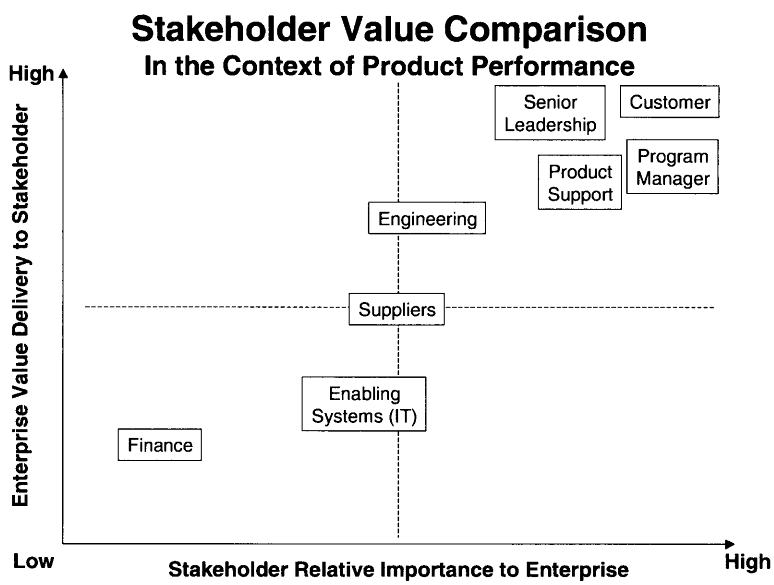

Figure 4.2: Case Study Stakeholder Value Comparison

Additionally, if any stakeholder within this enterprise wants to assess product performance, all data required to do so must come from the customers. A Jawahar and McLaughlin paper (2001) supports this statement. Jawahar takes a lifecycle approach and addresses which stakeholders are more important in various phases of a business organizational lifecycle. In the case of this product and the length of time it has been in operation with a supporting business team, the enterprise is considered to be in a mature stage of its organizational life cycle; that is, the enterprise considers itself successful, and its rate of growth has flattened out. To continue strong cash flow, the needs of most stakeholders will be dealt with in a proactive manner. Therefore, it is important to understand the needs of the customer, the stakeholder with the highest relative importance to the enterprise and to whom the enterprise develops the highest value, and to execute on a plan that helps address their needs. Customer needs are driven by the product performance metrics.

4.2.1: Metric Data

As mentioned earlier, most of the product community meets during periodic conferences to review performance data, and each customer presents a unique plot package of this information. Unfortunately, the data on each plot package varies significantly across the customers, as not only are the metrics different, but also are the template formats in which they are presented.

Analysis of the data shown during one recent customer meeting exemplified the issue of metric divergence. Over 180 individually named metrics were presented during this conference. Figure 4.3 below shows the amount of divergence across seven product customers by metric name only, as this could not be assessed at a lower level because definitions for each metric vary from customer to customer (another recognized issue). Therefore, perhaps the customers do track the same things in the same way, but they are interpreted differently on both a customer-to-customer and customer-to-developer basis during these conferences.

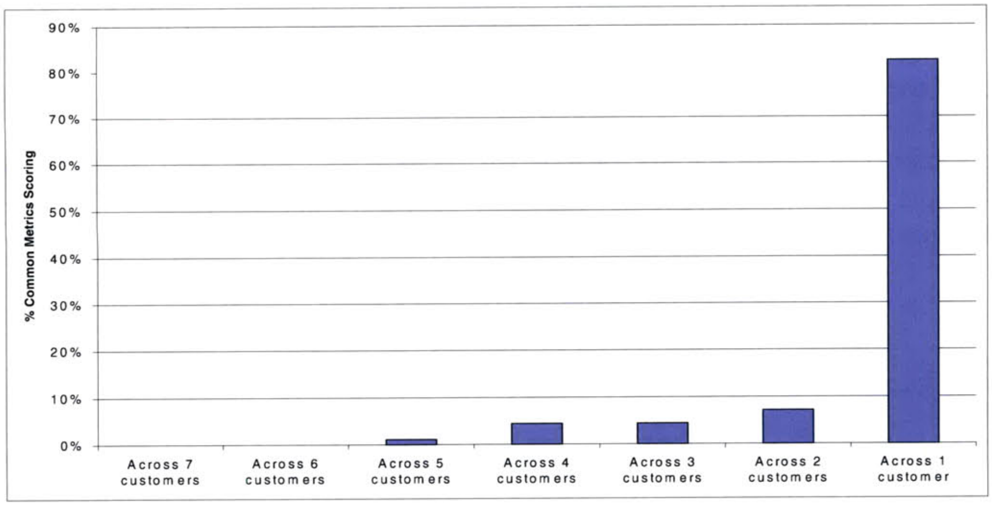

Figure 4.3: Percentage of Common Product Metrics as Presented by the Product Customers

Over 80% of the metrics presented are specific to one customer; thus, it is hard to establish a level of commonality between customers in an effort to improve the product across its entire platform. This is not to say that the product has not improved over time, because it *has* improved, but there is a growing frustration within the product community on how to interpret various customer metrics more holistically.

4.2.2: Customer Email Exchange Data

The five propositions from Yin's case study method listed above were rewritten in the form of ten open-ended questions that were asked to five of the product customer contacts. They focused on understanding the data systems that are used to input and output data, how the metrics evolved into the customer's operating environment and which stakeholders were involved in the metric creation process, how the metrics are used to drive business decisions, and how often the metrics changed during the product's operating life cycle. The ten (normalized) questions asked were:

- What is your [enterprise's] process of recording raw metric data (types or numbers of databases used)?
- How did the current [product] metric recording process come to fruition? How was it developed, how have things changed along the way, and how much influence into the system did you have vs. [the product developer's] influence?
- What are your methods of recording data (spreadsheets and databases vs. observations, field reps to record information, etc.)?
- How long have you been recording [product] metrics?
- How would [product] metric standardization benefit the way you run your [enterprise]? What improvements would you like to see currently in your system?
- What are your concerns in standardizing [product] metric data?
- What [product] metrics does your [enterprise] track outside of [product] data?
- Do your [product] metrics measure your key processes? What are your key processes?
- Are your strategic objectives driven by your [product] metrics? What are your strategic objectives?
- How often are [product] metrics assessed and re-evaluated?

The amount of information that was provided varied from customer to customer, but many of the propositions were supported by the responses to these questions, among others:

- Customers have multiple unique databases of which they input and output data. Not only are the databases unique, but the quality of the data also varies by customer. Additionally, in some cases, manual work is required to take product data and calculate or graph the results to be able to show product performance to others, thus increasing opportunities for human error. Some databases are reactive in nature; that is, they are created by product events that resulted in poorer product performance rather than recording data that proactively foreshadows adverse performance.

- Some customers have more experience than others in either using this product in its operating environment and thus are more knowledgeable in the way it performs, or have greater capabilities and support to record more data. Some customers believe the data recording system they use is adequate for their business needs as well as for the people who need to use the databases. Other customers recognize there is room for improvement. Some customer metric databases are created with a significant amount of product developer input, and some are not. Some are created based on the primary product customer's metrics as a starting point, and then made internal changes specific to their enterprise.
- There is also a "some things never change" mentality: it is harder to change the way an enterprise does business if that enterprise is used to recording data in a certain way, or has been thinking a certain way, for a significant amount of time. A recognized effort is necessary to change.
- Certainly a component to the above observation, but a barrier in and of itself, is leadership involvement. In some customer cases, a manager at the time of metric inception decided what metrics needed to be tracked. As leadership changes over time, so do the product metrics.
- In the case of this product, not all customers can share their data freely with other customers, so some recorded metrics are formed internally rather than with guidance

from the product developer. This leads to individual metrics for individual customers without the understanding as to why these metrics are being recorded.

- Customer metrics are re-evaluated at different intervals, ranging from monthly evaluations to updates every two years.

This data reinforced the propositions of why metrics have diverged across product customers in the first place, as well as defined barriers that would need to be worked through in order to commonalize metrics across all customers.

4.3: Identifying Metric Commonality Benefits

Much of the case study thus far has focused on the problem at hand; attention must now shift towards the solution. A research design, Figure 4.4, was created to help drive this issue to a solution.

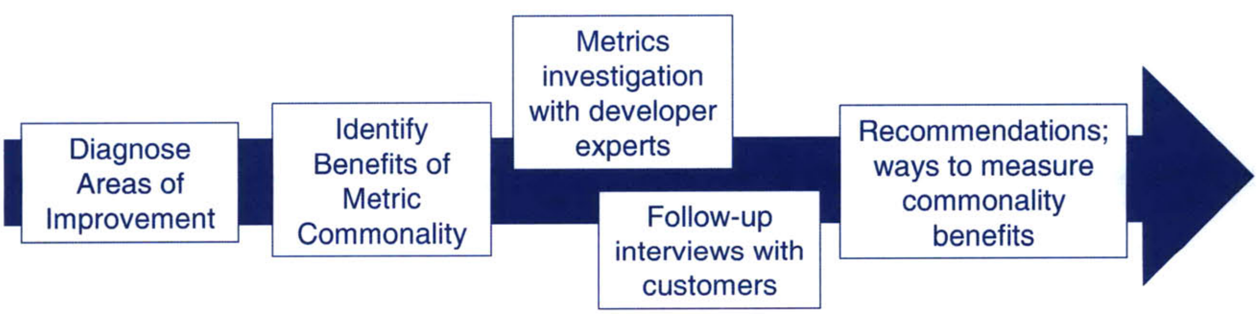

Figure 4.4: Research Design to Identify Common Metrics for Case Study

4.3.1: Diagnosing Areas of Improvement

The current state analysis of this enterprise can be summarized in a few bullets, which can be classified as the areas of improvement:

- Customers track different product performance metrics, due to a number of reasons: the use of various database systems with varying levels of data quality, varying levels of knowledge or capability (technological and technical) to record the right performance metrics, leadership involvement and historical aspects as a motivator to change (or not change) the current recording process, and the inability to share all performance data due to export control.
- Each tracked metric does not have a standard definition used by all customers.
- Product conference presentations are in varied presentation formats, making it more difficult to properly deduce the right information and compare against all customers.
- There are periodic reviews of performance metrics, but it is unclear if the metrics change because it is the proper thing to do given the product's place in its operating life cycle, if customer leadership instructed a change, or otherwise.

4.3.2: Identifying Benefits of Metric Commonality

A bottom-line benefit of the use of metric commonality for each of these current-state bullets was the necessary next step to begin focusing on a solution. There must be a motivation for change, and this connection acts as that link that would allow the customers to understand the importance of using metric commonality to benefit their enterprise. Before making these connections, the customers also addressed what they believed were the benefits and detriments of incorporating metric commonality into their enterprise. In general, the customers are supportive of metric commonality. The customers look forward to improved information sharing. All customers tracking the same data would result in comparable product statistics and metrics, and would be able to speak the same language, as everyone would be tracking an optimal set of metrics with the same, standard definitions. It would be less complicated to determine which customer has the best product performance and which customer has the worst product performance, in an effort to continuously satisfy the customers with improved product performance. Tracking, and interpreting, the same metrics will also allow the community to determine not only special causes that have an effect on product performance, but the community would also be able to tell if previous fixes to performance did, in fact, make a difference across the entire product operating line (i.e., the fix was value-added). If the data points to a specific fix to better the product, understanding which customers showed more of a performance problem would allow the community to better determine how to divide up the improvement money to fix performance. The success of this new decision-making tool would also have the potential to become a best-in-class initiative, and could be used across other product lines if proven successful.

Coupled with benefits was also a set of detriments. Again, as defined by the customers, one is resistance to change. If the customer has a metric recording system that they believe is sufficient for their enterprise's needs, they do not believe they should have to change the way they do business. The new way of recording data would need to be highly compatible with the current data recording process or system. And finally, there is the issue of the inability to share all information across all customers due to their enterprise's regulations, restrictions, and safety. These issues will be discussed later in this paper, as the methods of adoption will provide abatement plans for some of these detriments.

The coupling of the areas of improvement and the customer-defined benefits of metric commonality helped generate metric commonality benefits for each of the areas of improvement listed in Section 4.3.1, as shown in Table 4.1.

Table 4.1: Identifying Benefits of Metric Commonality for Each Area of Improvement

Improvement Opportunity	Benefit of Metric Commonality	Bottom Line
Customers track similar high-level metrics but use different measurement systems	Tracking the same high-level metrics will reduce variation in what is analyzed. Less variation in data means more accurate assessments of the data. Less time will be needed to interpret the data, as well as more clarity of what root causes drive the high-level metric behavior. Communication between customers will increase. Identification of the right corrective actions will be recommended.	Cost savings.
Each tracked metric does not have a common definition across all customers	Less metric variation and uncertainty reduction in data interpretation. Less time will be needed to interpret data.	Cost savings.
Conference presentations show varied metric information using varied presentation formats	If the same information and same formats are used, then less time and effort is needed to interpret the data presented. The communalization of what is presented will allow the customers to better share information between other customers, initiating a "best-in-class" work flow, as well as an increase in universal product knowledge. The program manager can also use this information better to determine what improvement programs should be implemented to improve the product's performance.	Performance improvement should decrease maintenance costs.
Lack of understanding in how metrics change over the course of the product's operation	Tracking the right metrics at the right time leads to a better understanding of product performance throughout its lifecycle, and therefore improvement money can be spent on the right programs at the right time.	Increased performance and decreased maintenance costs.

Qualitatively, the adoption of metric commonality for this product would result in cost savings for both the customers (the ones maintaining the products while in operation) as well as the product developer (the ones performing the engineering improvements that would improve the product's performance).

4.4: Recommended Performance Metrics and Adoption Attributes

4.4.1: Performance Metrics Investigation with Developer Experts

Separate to this research, but still considered a deliverable for the enterprise in this case study, was the creation of an optimal list of performance metrics that should be recorded, and shared, by all customers in an effort to streamline the performance metric interpretation process. This baseline was established through semi-structured interviews with product developers. Eleven design engineers with expertise spanning a number of the product's individual components, customer-interfacing engineers, safety engineers, and systems engineers were interviewed to help establish this baseline. These engineers were asked approximately a dozen questions, broken up into three sections. The full list of (normalized) questions is listed in Appendix B. The first portion of the questions helped to establish this baseline – the experts recommended what they think were the most important metrics that would address the voice of the customer across the product's various operational life cycle phases. Additionally, the experts recommended what performance metrics would need to be recorded to better allow the expert to perform his or her job in improving the product. The second portion of questions was more abstract in nature. The experts were first asked for their thoughts on how effective metric commonality would be across the product line in question, and were then questioned about what motivating factors and techniques they would employ to more easily convince the customers to adopt this new decision-making tool. Finally, in a separate communication, the experts were also asked to rank a number of researched adoption attributes in response to what characteristics they believe are most influential and value-added in changing the way decisions are made.

4.4.1.1: Expert-Recommended Metrics Analysis

These experts recommended just under one hundred metrics (a forty-five percent reduction than the number of metrics used in the historical data set) that addressed either the voice or customer, the data required for a thorough engineering analysis, or both. This list of metrics was then sectioned into four groups, depending on the frequency of the same response, as these four groups gleaned insight into which metrics the experts considered to be most value-added in making decisions from these performance metrics. As hypothesized, a small group of five metrics were recommended by more than half of those interviewed, and are therefore considered the most value-added metrics all customers should track. Five other metrics were recommended for tracking by more than twenty-five percent of those interviewed. Twelve metrics were recommended by at least two of the eleven respondents. The remaining seventy-eight percent of the metrics were recommended by only one expert, and were also mostly the metrics that would be value-added for a specific engineer in a specific function to better do his or her job in improving product performance.

Similarly, the experts were asked, for each recommended metric, when during the product's life cycle it would be an optimal time to record this information. The life cycle was divided into three sections: the product's entry into service, the phase when the product is considered in its prime operation (most mature), and the phase when the product begins to phase out. The results of this analysis are shown in Figure 4.5.

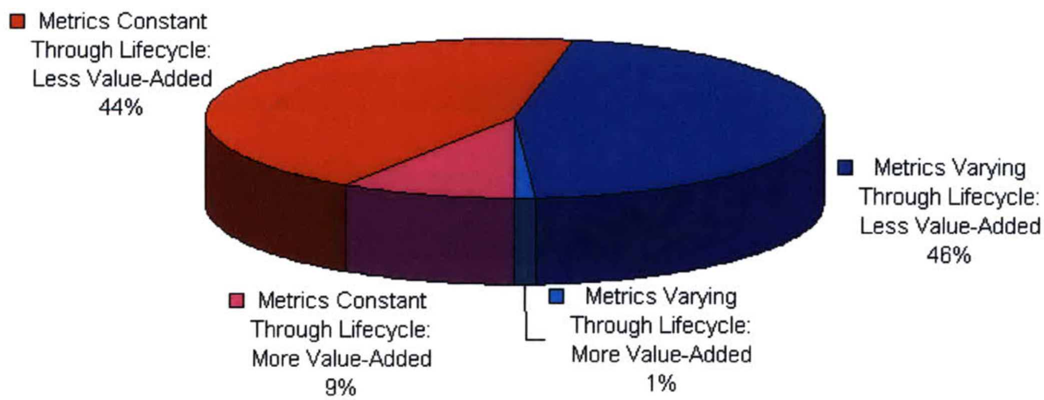

Figure 4.5: Common Metrics Staying the Same or Varying Throughout a Product's Operating Life Cycle

According to these experts, fifty-three percent of all recommended metrics should be recorded during all phases of the product's life cycle, and of these metrics, eighteen percent are recommended by over twenty-five percent of the experts that were questioned (nine percent overall). The remaining forty-seven percent of all recommended metrics were valued differently during different phases of the product's life cycle, and therefore do not always need to be tracked. Of these metrics, only two percent were recommended by over twenty-five percent of the experts that were questioned (one percent overall). Additional analysis showed that the ten percent of metrics that were considered to be the metrics of higher value in assessing product performance all addressed either the voice of the customer, or the combination of the voice of the customer and the expert. Most metrics that were specifically geared to help the product developer better perform his or her job were flagged as a less-valued metric that should be tracked, which is also consistent with the research hypothesis.

There are three significant takeaways from this analysis: (1) less than a dozen metrics would need to be tracked, as a starting point, across all customers for the product community to effectively understand product performance in its various operating environments; (2) metrics that address the voice of the customer provide a more effective assessment of product performance than metrics that are specific to the product developer; and (3) metrics that are recorded consistently across the product's operating life cycle provide a more effective assessment of product performance than metrics whose values vary throughout the product's life cycle.

4.4.1.2: Expert-Recommended Metric Commonality Effectivity and Motivation

The purpose of this section of the interview was a way to determine if those that develop the product agree that the product's performance can be assessed as effectively (if not more effectively) as it is now using a minimal amount of data. The experts were asked what they believe was the optimal percentage of common product performance metrics across all customers that would result in a maximum understanding of the product's operational performance. Overall, the responses varied (below is a sample):

- "From a safety perspective, the most effective measure is [Measure "X"]. This is a single metric that should be common across all customers, and is 80% of the entire story."
- "You want to be simple. The fewer metrics, the better. You tend to plateau. Currently – people invent their own way to measure metrics to look good, [which is] not the way it should be."
- "At least 6 high-level metrics with a lot of sub-tiers. Should be able to catch all the little faults. Looking at these 6 metrics, you can explain 80% of the performance of the product."

- "Hard to quantify individually because each discipline generally needs to have some basic metrics that they need to know about so they can be proactive in fixing things. It may only be a half dozen per area of expertise, for example."
- "At least 3 metrics [defined in earlier portion of the interview] are required across all customers. Don't want to water things down too much, but you need some sort of high-level quick snapshot of general trends. In terms of common metrics across all customers – between 50% and 75%."
- "Audience matters. When talking conference audience, most customers have reached the maturity phase requiring [Metric "Y"] so they are sharing a lot of the same issues. So it makes sense to try to get them to present things in similar ways in order to maximize good data sharing and leveraging for earliest identification of trends."
- "You need a certain amount of data to do anything, but once you do too much, you lose value. For example, no need to measure "Z" eight different ways; can you monitor "Z" in less than eight ways across all customers? Having one metric isn't enough, but having a few metrics related to one area may be enough to do a lot. If you get more data after that, it may be too much information and not be too helpful."
- "More than 6 primary metrics is too much, for program management."

While there is some disagreement between the experts in how many metrics should be commonalized, or the overall percentage of metrics that should be common across all customers, all these responses highlight that the experts believed the same information currently obtained during these product conferences could similarly be obtained using a significantly smaller number of metrics than those recorded and presented today (which is supported by the recommended metrics analysis in the previous section). This conclusion is an important second-order response for this research, as the experts, as well as the customers, would need to support this new decision-making tool for it to be successful.

Additionally, the experts were asked to envision themselves as a customer of the product. While maintaining this alternative viewpoint, they were asked three questions: (1) what data they would want to be shown to convince themselves that metric commonality is a more effective and accurate tool to make decisions; (2) the amount of customer and developer interaction that is needed (and when) to determine this optimum list of common metrics; and (3) if any incentives, or incentive structures, should be in place to help motivate the adoption of this idea. These were preliminary questions that helped address the concept of adoption.

A number of different ideas were presented to help answer question (1). Some responses were motivated by creating an individual customer's burning platform ("Something that would point out that the way I'm doing it today is not accurate," "Let the data speak for itself: easy to determine who is best in class and who is worst in class. The first step is knowing that you're different, and then you can take the initiative to find out why," "If there was a program/platform/customer that was low performing and you communicated that rate to the poor performing customer and compared it to the rest of the community, then the data may turnaround because nobody wants to be the worst performer"), while others motivated change by using an example of a product or program where commonality was employed successfully ("In an ideal world, you'd like to see results from some [enterprise] that has been successful with it"). Some responses highlighted the effectiveness of reducing metrics ("Here's what's lost when you chase too many metrics," "Just knowing that all metrics are the same across all customers would be sufficient. We monitor this, that, and the other thing so we can be proactive in fixing problems. That's the advantage with commonality – so we are all talking the same language," "Prioritize importance of metrics per customer and make sure that those are directly in the common metrics list"), while others made a connection to the bottom-line measures ("Life cycle costs and spare parts will be minimized. Fewer parts to be removed, less in the inventory," "Show a change in improvement program funding based on common data").

While there is variation in when, and how often, the product's customers and developers should interact in generating an optimal set of performance metrics, all experts responded positively that interaction must occur. Some experts responded that the customer should be involved in the generation of performance metrics from the inception of the program, as at the end of the day, the customer will be the one using the product and providing performance data back to the developer. Others believed that the developer should create a set of performance metrics first and then present the metrics to the customer and begin an iterative process. One respondent identified the use of an integrated product team: "Might need one person from each customer to contribute. Somewhat sequential teams. Internal team comes up with proposal, external team gains the consensus of the customers." Two respondents stressed the importance of creating a database system that would support the recording of the optimal list of performance metrics. "Presenting the examples of data systems that would help [the customers] realize what it takes to support [their enterprise] logistically and illustrate the advantages of commonality with the prime customer for technical support data systems that could be applied to a new customer," "The quality of data is also important to measure. Need to get the customer committed to doing this right off the bat."

The responses for the incentives question complemented and first question of this set, as the responses were less about incentives or incentive structures and more about the reasons why the customers should adopt this new decision-making tool. A majority of the responses wanted to show the customers the amount of money they would save in maintenance costs by using a common set of metrics. A couple responses stressed the anticipated improved communication within the product community, as commonalizing metrics would result in a lack of confusion or ambiguity in data analysis, and better understanding of product performance. One respondent identified one of the two top motivating factors of adoption: commonalizing metrics would result in an ease of use. Another respondent identified the model as an opportunity to redo the way business is currently done, as metric commonality would result in a more efficient, effective way to analyze the same data.

4.4.1.3: Expert-Recommended Adoption Attributes

Up to this point, the experts provided their thoughts on methods and strategies that could be used to convince the customers to change the way they record metrics. Also important was the understanding of adoption attributes that would be critical for its success. The experts were provided with a list of twelve attributes developed by Valerdi and Blackburn (2009), and were asked to rate each attribute in terms of three categories (assigned values of 3, 2, and 1, respectively, for data analysis):

- *Must Be*: referring to attributes where user is dissatisfied from its absence but never rises above neutral no matter how much of the attribute exists;
- *One-Dimensional*: referring to increasing user satisfaction from the presence of this attribute and decreasing satisfaction from its absence; and
- *Attractive*: indicates areas in which the user is more satisfied when the measurement system has the attribute but is not dissatisfied when it is absent; lack of an attribute leads to a neutral reaction.

The survey and definitions of the 12 attributes are listed in Appendix C. Results of the survey are shown in Figure 4.6.

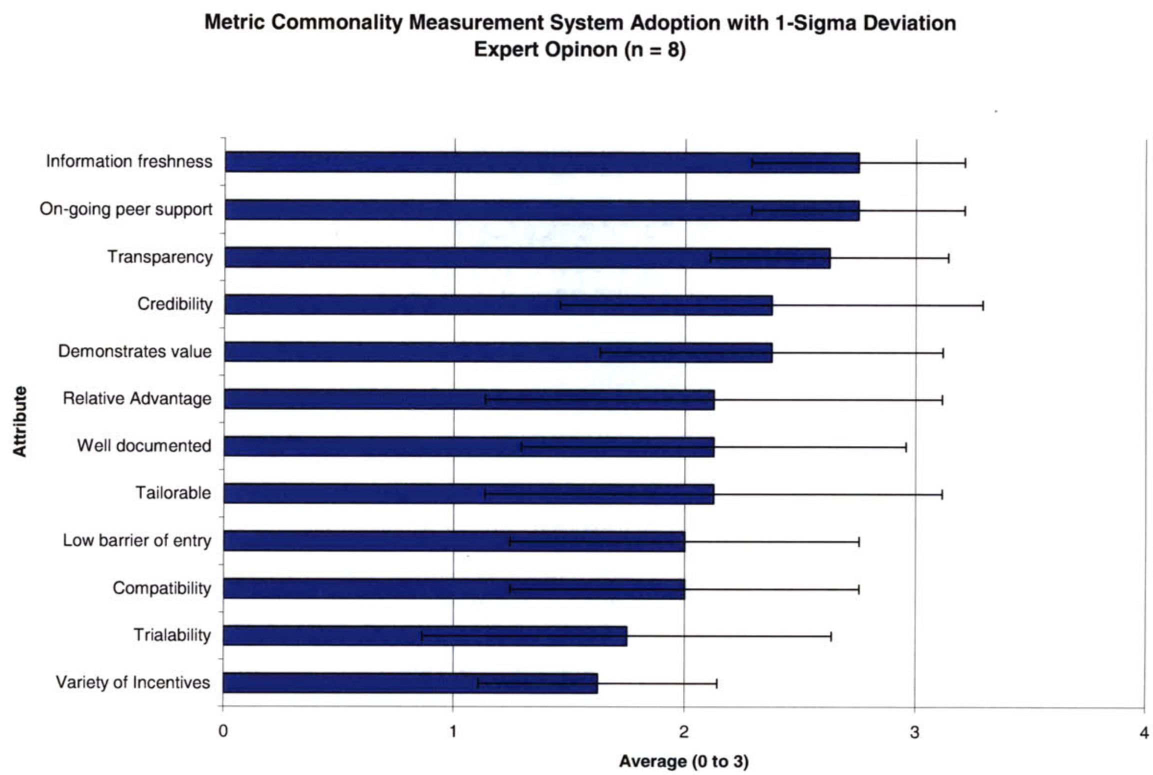

Figure 4.6: Expert-Defined Metric Commonality Measurement System Adoption Attributes with 1-Sigma Deviation (n = 8)

Experts believed the three most important attributes for adoption of a metric commonality measurement system were information freshness, on-going peer support, and transparency. These three attributes tie directly together to the capability of the database system that would be created to record metrics, as well as the quality of the data. They believed the measurement system should be supported capability-wise by IT experts if the database of information is IT-enabled, and should be supported by knowledgeable field representatives in terms of what data is placed into the system. The recorded data should be updated on a periodic basis in order to keep performance current and relevant for making decisions. The system should act as a "pull" system so the product community has easy access to this data.

The experts identified the three least important attributes for adoption of a metric commonality measurement system as compatability, trialability, and variety of incentives. Compatability and trialability speak to a complete revamping of the current metric recording system; the experts did not believe the measurement system needed to be compatible with the enterprise's current operating environment, nor did they believe that the measurement system should be piloted. If a new, better system can be created, then it should be enforced. Additionally, they did not believe the measurement system should include personal incentives; it should be a standard measurement system for all customers to input data.

While the attribute survey data appears to show a gradual trend in the data, there are statistically significant differences within the data set. A one-way ANOVA test shows that the top three attributes are statistically significant from the remaining nine attributes ($p=0.00$, $\alpha=0.05$), and the bottom three attributes are statistically significant from the remaining nine attributes ($p=0.03$, $\alpha=0.05$). Tracking each metric cluster individually yields a marginal statistical difference result ($p=0.058$, $\alpha=0.05$), implying that not all attribute rankings could be used from the same metric set.

4.4.2: Follow-Up Interviews with Product Customers

Semi-structured interviews with four of the product customers followed. These four customers represented approximately 80% of the product population and, as such, all have similar typology. The objectives for this set of interviews were threefold. They intended to:

- Probe into the value of the current recorded metrics with regard to the customer's enterprise;
- "Close the gap" between the expert-recommended performance metrics and the metrics each individual customer tracks and shares publicly in terms of product knowledge, recording capability, or something else; and
- Address the attributes that are necessary for the customer to adopt metric commonality as a new decision-making tool. The results are also compared to the results from the expert interviews.

The customer interview questions are listed in Appendix D. Note that the questions have been normalized.

4.4.2.1: Further Insight into the Current State of the Program

Despite the lack of common metrics for this product across all customers, the subset of customers that were interviewed all identified the same goal they currently try to obtain for their individual enterprises. This provides a good starting point for common metric determination, as these customers are trying to achieve the same thing. The differences in determining customer metrics, therefore, begin at some lower level than the overall goal.

Referring to Basili et al's "Goal-Question-Metric" approach (1994) to determine the right metrics, customers all have the same goal, or conceptual level, which is to maximize the amount of time this product is in operation without having to perform maintenance on it. The differences either begin during the question creation to characterize how the goal is going to be performed, or during the metric creation that would help answer each question quantitatively. Conclusions from this section of the customer interview could not pinpoint which of these two steps provided the divergence, as the procedure of metric creation varies across the different customers. It therefore is plausible that metric divergence occurs during both the question creation and metric creation parts of this overall recommended method.

The primary product customer, also the customer with the most storied history of operating this product, discussed its metric creation process. There was a significant shift in what metrics should be recorded approximately ten years after the product was already in operation; it was at this time that the most value-added metrics for this customer's enterprise were created. The metrics were created with an internal integrated product team, with some influence from the developer's field service representatives. However, as stated from this interviewee, "we always listened to inputs, but we didn't necessarily agree with inputs."

The next largest customer had a different response to this question. This customer said that "a good part of [the metric creation] was by experience and collecting all the data over the years." There was also some influence from the developer engineers as well as other stakeholders that participated in these periodic product conferences. However, it was (and still is) the role of the person interviewed to ultimately approve or deny which metrics should be tracked.

A third customer had a very different creation process. Most of the metrics currently used for this customer were taken from historical data, with updates and tweaks throughout time. This customer utilized metrics that were presented by other customers during these periodic conferences and "tried to make the best mix and have the metrics that make the most sense for me." This customer noticed, in this process, that not all customers were tracking metrics in the same way, but tried to take what he considered the "best ideas" and utilize them for his own enterprise. In this customer's opinion, "the [primary customer] metrics are not clear and it is difficult to see [primary customer] trends, so I tried to make things simpler" by referencing metrics that are used by other customers. In this situation, the product developer did not help with the metric creation process.

The fourth customer interviewed assumed the performance metrics role with a set of metrics that were already in place from the metrics predecessor. This metric list was a combination of integrating the list with other similar product platforms within the customer's enterprise, as well as using the primary customer's metric list as a starting point. The interviewee also suspected that there was "probably [conference] influence" as well.

What is common between these four customers is the goal, or, the reason why metrics are being tracked in the first place. However, the methodology used by each customer to generate this list of metrics that would help drive decisions towards this common goal is unique for each customer. What must be changed, then, is the development of the questions, and corresponding metrics, that address this common goal.

4.4.2.2: Comparison of Customer-Generated Metric List to the Historical and Expert-Recommended Metric Lists

These customers were asked what five to ten metrics they considered most important in addressing their enterprise's values. This generated list currently serves a number of purposes: (1) it provides insight into a "wish list" of metrics for each individual customer, (2) it identifies the level of commonality in metrics on a customer-to-customer basis, (3) it identifies the level of commonality in metrics between customers and the expert-generated metric list, and (4) it either supports or does not support the use of metric commonality as a decision-making tool for this particular case study.

Purposes (1) and (2) are addressed together. In total, the four customer contacts that were interviewed generated twenty-eight metrics that they considered most valuable for enterprise success. Of these twenty-eight metrics, one metric was common across all four customers, two metrics were common across three customers, five metrics were common across two customers, and nineteen metrics were common across a single customer. Similar to the expert-generated data, each customer had a "wish list" of metrics to track, and almost thirty percent of the total customer-generated metrics were common across at least a quarter of those interviewed. This trend is similar to the trend found when the experts generated their recommended set of metrics (ten percent of expert-generated metrics were common across at least a quarter of those interviewed), though not as selective as there were more experts interviewed than there were customers interviewed. Nonetheless, there is some level of commonality that exists between these four customers.

Figure 4.7 helps address purpose (3). This figure shows the amount of common metrics between the three data sets obtained throughout this case study. Three sets of data helped generate the values in this table; the top ten historical common metrics (as mentioned in Section 4.2.1), the top ten expert-generated value-added metrics for the product program, and the customer-generated metrics (n=28).

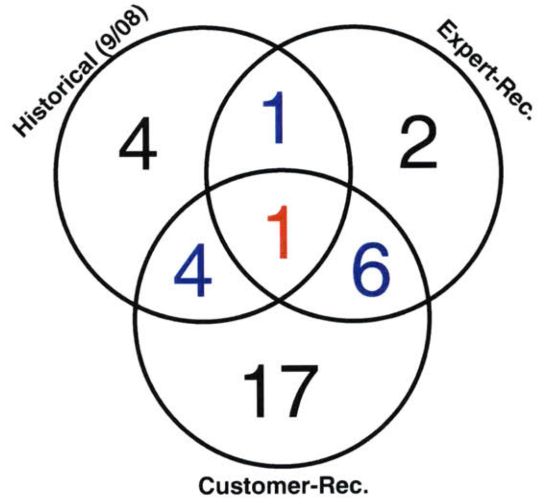

Figure 4.7: Venn Diagram of Shared Metrics Between Historical, Expert, and Customer Data

It is understood, from the current state analysis, that there are few metrics that are common across the historical conference presentations. This is reflected in the Venn Diagram of only showing one common metric between the combination of the expert-generated and customer-generated metric lists. The new direction of the program would generate metrics from some combination of the expert-generated and customer-generated metric lists, and the diagram sows there is more commonality between these two metric sets than any other combination of metric sets. Overall, twenty-three percent of the expert- and customer-generated metric lists are common between the two sets, a higher number than the single common metric between all three metric sets.

While the empirical data that addresses purpose (3) begins to address the utilization of metric commonality for this case study, the results in Table 4.2 further drive this idea. Table 4.2 helps address purpose (4), that metric commonality can be used as a decision-making tool for this particular case study.

Table 4.2: Summary of Historical, Expert, and Customer Data Results

		Metric Population Set			
		Historical Expert Customer	Historical Expert	Historical Customer	Expert Customer
Overall	Total Number of Metrics in Set	35	18	33	31
	Number of Shared Metrics	1	2	5	7
	Percentage	3%	11%	15%	23%
Number of Metrics in Agreement with "X" Number of Customers	0 Customers	0	1	0	0
	1 Customer	1	1	4	4
	2 Customers	0	0	1	1
	3 Customers	0	0	0	1
	4 Customers	0	0	0	1

"goodness"

Table 4.2 can be broken down into two sections that help support purpose (4) of this study. The first section of data (labeled "overall") identifies the number of shared metrics within certain data sets. Only three percent of all metrics are common across all three data sets as the current state of the program addresses a lack of commonality in historical data. As expected, the percentage of common metrics increases when the historical is data is compared to the expert-generated data (eleven percent), as some of the expert-generated metrics were defined to address the voice of the customer, assumed to be the sort of data tracked in the historical metric set. The combination of the historical data with the customer-generated data yielded a higher result of sixteen percent. This is in agreement with the case study as well because the customers created the historical data. It is therefore expected that there should be more commonality between the historical data / customer-generated data set as compared to the historical data / expert-generated data set. If one neglects the historical data as a set and focuses on common metrics between the expert-generated and customer-generated lists, the most amount of commonality exists (nearly one quarter) of all recommended metrics.

Similarly, the second section of data (labeled "Number of Metrics in Agreement with 'X' Number of Customers") also supports the proposition that metric commonality can exist within this case study. Identified in the columns for this section are the number of metrics that are shared across 'X' number of customers, defined in the section rows. What is considered "goodness" is the number of common metrics between 'X' customers across the four different metric subsets increasing as the amount of commonality within that subset also increasing. A proportional increase would identify the potential existence of metric commonality within an enterprise, and a non-proportional increase would identify the elimination of metric commonality within an enterprise. The results of this case study yielded the potential existence of metric commonality, therefore, proposal (4) supports the hypothesis of this thesis.

Also to note are potentially uncommon metrics that are actually the same thing, but use a different metric name. For example, the definition of the primary customer's most significant metric (Metric X) is, in fact, the same definition for the other interviewed customers with a different name. The second largest customer stated that Metric X "is a [primary customer] metric, and sounds a lot like our [Metric Y]." Similarly, the third customer interviewee stated that Metric X "is the same as our [Metric Z]." Therefore, if all customers are using the same definition of a recommended metric, then in this example, Metric X = Metric Y = Metric Z, another common metric. Therefore, there may be more commonality than what is currently analyzed, but this will not be known until all definitions of all metrics have been reviewed as well.

Based on these results, the result of purpose (4) is that there is potential for this case study enterprise to use common metrics across all product customers; what has not yet been performed is a quantitative example of the benefit from doing so.

4.4.2.3: Customer Adoption Attributes

The same adoption survey that was provided to the experts and referenced in Section 4.4.1.3 was presented to the customers. Figure 4.8 represents the customer responses compared to the expert responses.

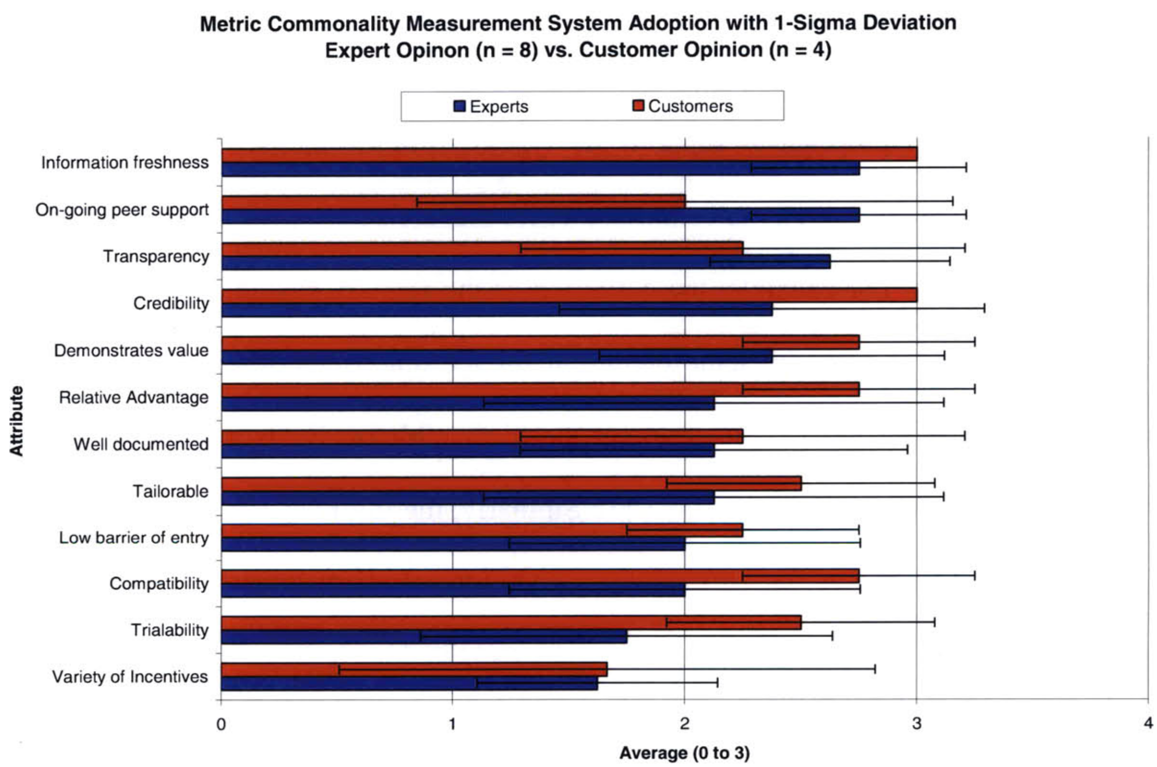

Figure 4.8: Expert-Defined and Customer-Defined Metric Commonality Measurement System Adoption Attributes with 1-Sigma Deviation (expert n = 8 and customer n=4)

There were two adoption attributes the four customers all ranked as "must-be:" information freshness and credibility. Customers want the metric data to be updated on a periodic basis so the data can help them make the right decisions in how to allocate their money to improve the product. Additionally, the customers do not want to change the way they do business unless they are presented with an example of this decision-making tool working successfully for a different program.

The three least important attributes for adoption of a metric commonality measurement system were variety of incentives, ongoing peer support, and low barrier of entry. This new decision-making tool should not include incentives (and one interviewee refused to rate this attribute) as the model should be adopted knowing that the customer enterprise will improve the way they make decisions, and should not be dependent on obtaining "bonuses" for implementing such a system. Some customers believed that the system should act as a standalone system and support would not be necessary as all metric definitions would be predetermined prior to the adoption of the new tool; therefore, peer support is not critical for the success of the program. Further work would need to assess the reason why low barrier of entry was rated in the bottom three.

Statistically, customer data appeared to have a smaller overall mean differences and standard deviation compared to the expert-generated data. A one-way ANOVA test showed that the customer rankings of all twelve attributes are not statistically significant ($p=0.31$, $\alpha=0.05$). However, the breakup of attribute data in terms of top-three and bottom-three attributes yield different results. The top two attributes are statistically significant from the remaining nine attributes ($p=0.03$, $\alpha=0.05$), and the bottom three attributes are statistically significant from the remaining nine attributes ($p=0.01$, $\alpha=0.05$).

It should also be noted that there are differences between the customers and experts in terms of adopting this new tool, as highlighted in Table 4.3.

Table 4.3: Differences in Expert-Generated and Customer-Generated Adoption Attributes

Adoption Attribute	Expert	Customer
Top Three	Information Freshness	Information Freshness
	On-going Peer Support	Credibility
	Transparency	
Bottom Three	Variety of Incentives	Variety of Incentives
	Trialability	On-going Peer Support
	Compatability	Low Barrier of Entry

Ideally, the top three attributes and bottom three attributes should be similar between the expert and the customer, emphasizing alignment between these two stakeholders. Some of this is evident in Table 4.3: both sets of stakeholders agree that information freshness is a key attribute in adopting this new tool, and both sets of stakeholders agree that it is not necessarily that this model provide incentives with its integration into the enterprise. The overall conclusion made from this set of data, however, is that one stakeholder cannot assume what another stakeholders considers to be important (or not important), so decisions should not be made or acted upon without consent of all stakeholders. The adoption of a new decision-making tool is a significant undertaking, so it is important that both stakeholders address each other's concerns prior to its implementation.

4.5: Case Study Recommendations and Conclusions

4.5.1: Product Community Recommendations for Success

These recommendations will not only allow the product community to achieve its goals of efficient and effective product management and communication, but these recommendations also help overcome some identified barriers that were deduced from customer interviews. Note that with these recommendations come additional challenges to implement, but the focus of this case study is to recommend improvements to the way business is currently done.

Recommendation 1: obtain leadership support to both agree with the metric commonality measurement system, and also to help adopt it throughout all product customers. Leadership is a critical component of enterprise success. The Lean Advancement Initiative lab at MIT identified six value attributes from over one hundred candidates evident in successful programs, with leadership and management being one of the six (Murman et al, 2002). The benefits of metric commonality discussed in Section 4.3.2 correlate to the bottom line, the financial aspect, of any enterprise. Once the proper leadership understands the value that they could be receiving with this more effective measurement system, they may want to adopt it into their enterprise, and would initiate a transformation plan to do so. This recommendation also provides a plan to overcome two barriers identified in Section 4.2.2: varied leadership involvement and the historical component of "some things never change."

Recommendation 2: generate a common list of performance metrics, equipped with standard definitions, through an integrated product team (IPT). This study alludes to a list of less than one dozen metrics that, at a minimum, should be recorded by all customers in order to best understand, and make decisions to improve, product performance. What did not occur during this study was the joining of the customers, as well as the experts, in the same room discussing if these metrics are the metrics that everyone can agree to measure and from which to base decisions. Additionally, these metrics are just metric *names*; it is essential that each one of these metrics has a standard definition, agreed to by the IPT, that all customers will have the capability to measure. It is at this point where this product community can begin to compare "apples to apples." While this does not necessarily address the varied databases each customer maintains, it does address the barrier of the kinds of data should be input and output into these databases. Discussions can begin with the metric lists created from this research, and utilize Basili's goal-question-metric approach in determining the appropriate metrics (and metric elements) to address a common goal. Similarly, a metric "audit," as discussed in chapter two, should also be performed to identify metric alignment, false alarms, and gaps.

Recommendation 3: create an IT-enabled "pull" system of all common customer data, with appropriate IT support. The top adoption attributes as recommended by the experts are the importance of a database from which all information would be transparent, and accessible, to the entire product community. Such a database would enable better information sharing for customer-to-customer and customer-to-developer relationships, and would also allow for periodic data updates in an effort to keep the data fresh (another recommended adoption attribute among both customers and experts).

Recommendation 4: all customers should use the same presentation template during these periodic product conferences. Conference presentations currently show different metrics per customer. Presenting the same information on the same presentation template to the product community results in less time and effort necessary to interpret the data across all customers. This improves customer-to-customer and customer-to-developer communication, and the success of such a model can become a "best in class" initiative for other product lines to emulate. This recommendation may take the least amount of time compared to the other three, so perhaps this should be a first step in solving the issue rather than one of the last.

Recommendation 5: be patient, be energetic, and be supportive. Remember this process does not occur overnight. Quantitative results would start to be seen, at a minimum, after one to two years. The key here is to not lose faith in the transformation process. Always maintain the right leadership to support the process, understand the expected outcomes, and continuously engage the right stakeholders.

4.5.2: Case Study Conclusions

This case study is an example of what a number of enterprises inadvertently do: generate more metrics than necessary to effectively determine how well something is performing. For this case study in particular, these metrics are created by external stakeholders rather than internal ones, which makes program management harder. By analyzing this case study, insights are gathered as to why the number of product performance metrics has continued to increase over this product's operational lifecycle. Interviews with the product customers and developer experts have gleaned that, at the very least, commonalizing metrics across the entire product community can both exist as well as help streamline the length of time it takes to interpret data and draw proper conclusions. And while the number of metrics that should remain constant throughout the product lifecycle is almost split in half with the number of metrics that should vary through the lifecycle, ninety percent of all expert-generated "highly valuable" common metrics should be recorded throughout the lifecycle. Considering the right adoption attributes to use in a transformation plan, this product community shows great promise in benefiting from metric commonality as a new measurement system.

CHAPTER 5

CASE STUDY INSIGHTS, REFLECTIONS, AND IMPACTS

The beauty of case study research is the many insights that are formed during the research while attempting to answer the original research questions. While the case study addressed the research questions posed in Chapter 1 (which will be answered in Chapter 6), it also provided reflection on other topics of which enterprises can relate. The case study addressed the importance of the voice of customer, and the assumptions that enterprises may make in addressing this external voice. On a similar tangent, a part of understanding the voice of the customer is also learning how the customer makes decisions based the metrics they track. The case study in this thesis focused on the creation of a similar set of metrics for two decision-making stakeholders: the customers, and the internal organization. A section of this chapter will address the link between metrics and strategic planning for these two stakeholders, as the decisions that are made should be based on strategic planning. The case study also presented a new research design in order to update an existing performance management system that uses external inputs. While this framework is not perfect, it does provide a new way of addressing this issue, and its steps and limitations will be reviewed here. Finally, the case study results can be generalized to show its adaptability in addressing similar questions in contexts different than its own.

5.1: The Importance of the Voice of the Customer

This case study emphasizes the importance of not making assumptions about customer wants and needs. This is inferred from the following points:

- The metric commonality between the expert-generated metric list and the new customer-generated metric list is 12% more common than the common metrics between the expert-generated metric list and the historical customer-generated metric list. While this increased number corresponds to more commonality between the recommended metrics by the customers and the experts, the entire list still shows approximately 75% of differences.
- While there is some agreement in the most attractive and least attractive adoption attributes identified by the customers and the experts, there are still differences between the two lists, as was shown in Figure 4.8. What this means is the expert community cannot assume they know the best ways of having their customers adopt a new decision-making tool without working through the adoption process with the customers.
- One of the case study recommendations is to generate an integrated product team that uses a combination of experts as well as customer representatives to help generate definitions for a common set of metrics, in an effort to have all key stakeholders speaking the same language. The argument is that speaking the same language will result in more accurate and efficient decision-making in deciding upon the right product improvement programs to put in place.

In order to properly address the voice of the customer throughout this metric creation process, the customer should be involved in both the creation of the optimal common metric set as well as the most effective ways in adopting this concept as a new decision-making tool. In essence, the experts and customers would need to work together, in an iterative fashion, throughout this process. This concept of working together can also be called the co-creation of the voice of the customer, identified by Jaworski and Kholi (2006). They stress the importance of both the firm (similar to the case study experts) and the customers engaging about each other's wants and needs:

> *"In this co-creation process, the firm and the customers do the asking, listening, observing, and experimenting: that is, the firm and the customers engage in learning. The subject of study is customer needs/wants and firm needs/wants. The process results in the firm and customers knowing more about the needs/wants of the customer and the firm. Finally, after the process is complete, the firm and the customers figure out the goods and services that will be developed (or performed) by the firm and those that will be developed (or performed) by the customers."*

Also addressed in this literature excerpt is the recent revelation that customers may not necessarily know what they want or what would satisfy their needs or, if customers do know what they would like, they have trouble discussing it with the firm. This is also why understanding the necessity of mutual learning is also important. This was addressed in Section 4.4.1.2. Experts were asked how involved the customer should be in the development of these common metrics and, while the responses were varied, all experts agreed that the customers should be involved. Some recommended involvement from the beginning at the beginning of the metric creation process, while others recommended involvement towards the end, but all noted that customer involvement should result in an iterative process that drives in each stakeholder understanding the wants and needs of the other stakeholder. This also agrees with Jaworksi and Kholi who wrote that co-creating the voice of the customer requires "an open dialog between the firm and the customer. It requires a conversation over many periods of time, each time adjusting both the focus and mode of inquiry as the firm and the customer learn more about each other's requirements and capabilities."

5.2: The Link Between Metrics and Strategic Planning

The case study identifies the number of metrics that are most common between the customers and the experts and the expected efficiency and effectiveness in using these metrics to drive program-level decisions. It is hypothesized that these common metrics will help drive more efficient and more effective program-level decisions. What is not explored, however, is the relationship between these metrics and strategic planning, and making sure that the decisions the metrics help drive are the right decisions for the enterprise, based on the enterprise's strategic objectives. Table 5.1 reviews the common metric findings in Chapter 4.

Table 5.1: Common Metric Findings from Case Study

Metric Set	Total # Metrics	Top # Metrics	"Top Metric" Criteria	Remaining # Metrics
Historical Data	181	10	More than 50% customer agreement	171
Expert-Generated	99	10	More than 25% expert agreement	89

** Note: customer-generated metric set data is not included in this table as the triangulation analysis included all 28 customer-generated metrics, as each customer-generated metric set was already considered a top metric for that specific enterprise. If a similar "top metric" criteria is applied to this data set, then more than 25% of customers agree with 8 metrics, leaving 20 remaining metrics in the last column.

In attempting to link metrics and decision-making, Hubbard (2007) recommends asking the following five questions to help put proposed measurements in the right context. They are:

(1) What is the decision this is supposed to support?

(2) What really is the thing being measured?

(3) Why does this thing matter to the decision being asked?

(4) What do you know about it now?

(5) What is the value to measuring it further?

Hubbard breaks down the five questions as follows: the first three questions define what the measurement is within the framework of what decisions depend on its quantitative value. The last two questions should be asked before designing a particular measurement method.

Perhaps, in the "Remaining Number of Metrics" column, there exists metrics that should have a higher priority than common metrics because they help drive the right decisions for the stakeholder that needs them, based on the stakeholder's strategic objectives. Therefore, these "onesies" and "twosies" should not be ignored until they can be identified as alignment, false alarms, or gaps in the metric creation process between customers and experts.

5.3: A Framework for Updating a Performance Measurement System with External Input

This study creates a new research design and framework to address the research questions identified in Chapter 1. The framework follows the procedure outlined in Figure 5.1.

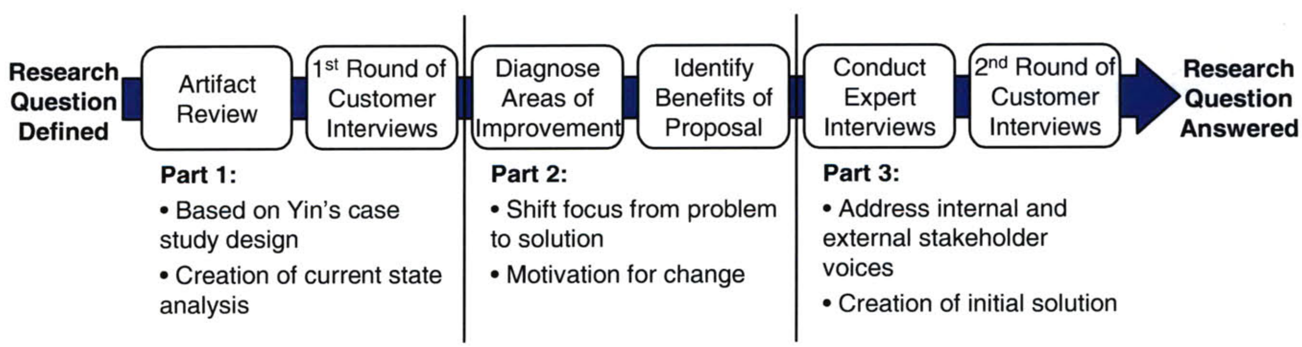

Figure 5.1: Case Study Framework to Create Performance Measurement System

The creation of a new performance measurement framework begins in part 1. While the intent of the first part of the framework was to identify the current state of the program and how it got to where it was today, the artifact review helped to identify the current metrics that are tracked across customers today because, as mentioned in Chapter 4, internal program decisions are based on external customer data. The artifacts used were the presentations from the other customers during these periodic product program conferences. The list of metrics was aggregated to form an all-inclusive list of 181 metrics. Now there is a starting point for metric collection before input from the key stakeholders was needed.

The second part of the framework is the identification of improvement opportunities, and the qualitative benefits of the application of metric commonality for the case study. This is the "ammunition" that is used for the third part of the framework, as it is easier for the key stakeholders to identify with the intent of the research if they are aware of the benefits of the research specific to their enterprise.

The third part of the framework is, essentially, a pre-audit. That is, the metrics that align between stakeholders, are considered false alarms, and are considered gaps, are identified on an individual stakeholder basis. Because the case study customers are global, it is difficult to have all customers in the same room at the same time, nor be in the same room at the same time with the experts. Therefore, all interview data collected from the customers (and the experts) was obtained via individual interviews. Perhaps this is a better system than including all stakeholders in the same meeting as this new decision-making tool was being explored, as this resulted in a lack of intimidation or judgment from other interviewees. Additionally, the anonymity in the data recording process allowed the interviewee to speak freely, knowing that his or her words would not be specified for anyone else participating in this study.

The experts and customers were both asked what metrics they considered valuable for decision-making; the resulting data from each interviewee was then combined with the data from other interviewees within the same stakeholder set. The data was then triangulated with the historical data that was obtained in part one of this framework. This produced a list of metrics that aligned (common between all three sets of data, or common between the expert-generated and customer-generated metric lists), that became potential false alarms (unique for the historical data set), and that became potential gaps (unique expert-generated customer-generated metrics, metrics that are common between the historical and expert-generated lists, and the metrics that are common between the historical and customer-generated lists).

It is at this point that the audit portion of the Medori and Steeple (2000) framework can be utilized, as now it would be necessary to gather all key stakeholders in the same room to review these three sets of lists; come to a consensus on the classification of the metrics being either aligned, false alarms, or gaps; and use the metric selection template (or similar) as defined in Table 2.1 to begin to commonalize, and standardize, the optimal metric set.

5.4: Beyond the Case Study – How One Set of Results Can Have a Large Impact

While the context of the case study is specific for a particular enterprise and its boundaries, the results can be generalized for other enterprises specifically with similar situations in managing and making decisions on metrics. In addition to another framework to update performance measurement systems as discussed in Section 5.3, there is also a connection to general lean thinking, a recognition of the importance of adoption, and a look into new strategies to perform better business.

5.4.1: The Connection to Lean Thinking

The Lean Advancement Initiative at MIT has pioneered the academic research of lean thinking for enterprises and enterprise transformation, as the LAI mission is to "enable the focused and accelerated transformation of complex enterprises through collaborative stakeholder engagement in developing and institutionalizing principles, processes, behaviors, and tools for enterprise excellence (LAI website)." Originally founded as a research collaboration with the United States Air Force to see if lean could be utilized in assisting military aircraft production, the LAI has been able to broaden their research to encompass enterprises across a number of domains that include, but are not limited to, the aerospace industry, military, and healthcare. As such, the LAI research focus has broadened as well, associating its research with four general questions that all enterprises ask themselves when attempting to transform their business in a more efficient and effective way, shown in Figure 5.2.

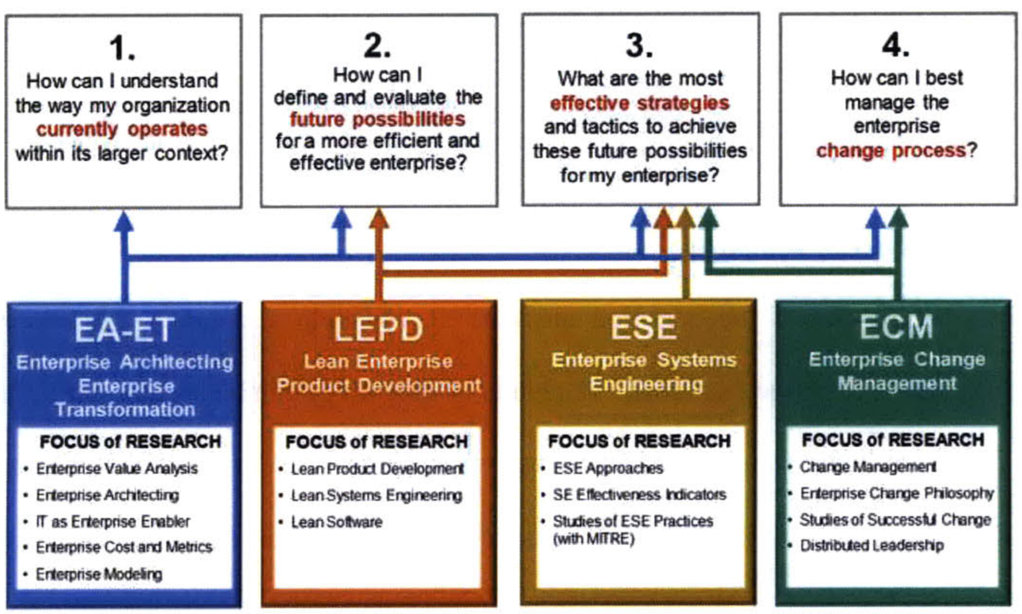

Figure 5.2: LAI's Four Research Questions

The research in this thesis addresses all four of these questions, and perhaps the steps followed in this case study can be used in other situations as well. The artifact review and initial set of customer interviews help answer the first question and begin to shape the current state of this case study program. Enterprise architecting, as well as enterprise metrics, is developed to help understand how the program developed over time to become what it is today. The bulk of the case study research addresses both the second and as well as the third question. The diagnosis of improvement opportunities and associated benefits of metric commonality serves as the future possibility of a more efficient and effective enterprise. The expert interviews, as well as the second round of customer interviews, address the effective strategies and tactics to achieve the future possibilities as these two stakeholders are considered the key stakeholders in this case study, and their interview results serve as the framework as an enterprise systems engineering approach to improving the current state of the program. The adoption principles that are also addressed in both these sets of interviews help to answer question four, as the adoption attributes address what stakeholders consider to be important in changing the way their enterprise does business.

5.4.2: The Importance of Adoption

While the results of the adoption attribute survey help strengthen the importance of understanding the voice of the customer, the results of the survey can be used as a mechanism for change. The right understanding of knowing what matters to key stakeholders can allow adoption attributes to be utilized as a persuasive tactic to influence and drive change in an organization.

5.4.3: Business Strategy Creation

Two business strategies resulted from the research of this case study. Both strategies focus on improved marketing in selling of the product to potential buyers.

The first strategy is to sell the product to potential buyers with an "optimal" list of optimal metrics. Certainly, it is important that the product meets the needs of its customers, but consider a customer that has never bought one of these products before. If the product developer goes to this customer stressing the product's applicability and practicality in this potential customer's environment, and also provide with the product an optimal list of metrics that will directly correlate to overall performance and drive the right kinds of decisions in further product improvements, this customer may find the product more attractive because, in essence, the work if determining what metrics should be tracked has already been solved, at least at a level of commonality of all the other current customers. This does not mean that unique metrics could not be created for this potential customer, but at least in this situation, the customer would already have a starting point.

The second strategy would be to offer the metrics to the customer as part of a "remote diagnostic" package. Perhaps the customer is only concerned in operating the product and does not care as much about the root causes of why the product operates the way it does. A remote diagnostic package, put together and monitored by the product developer, would therefore address both of these problems: the customer still does not need to use its resources in tracking product performance, while the product developer would have access to customer data in an effort to continuously make decisions on how to improve the product.

CHAPTER 6

CONCLUSIONS AND FUTURE RESEARCH

This research began with a concept, and ended with a new decision-making tool. This research proposed the marriage of metrics and commonality through the course of an operating life cycle, and investigated its applicability in a case study of an enterprise that makes program-level decisions based on external, customer-generated data. Through historical data analysis, expert interviews, and two rounds of customer interviews, the data gathered not only pieced together the current state of the program, but also proposed the benefits of commonalizing metrics specific to this program, and provided insight as to an optimal common metric list that should be utilized by all customers to better drive these program decisions. Outside of answering the research questions posed in Chapter 1, the case study also provided other insights addressing the importance of the voice of the customer, the link between metrics and decision-making, a general framework for updating an already-existing performance measurement system, and generalizability of the results.

6.1: Conclusions

This thesis began with three research questions. They have all been answered, to some extent, through this research.

(1) Can the concept of commonality be applied to metrics?

Yes. The results in Table 4.2 best represent the answer to this question. At the beginning of this case study, historical data showed 181 metrics across customers shown at one time during these product conferences in order to drive internal decisions on product improvements. After interviews with product developer experts, this list was nearly cut in half with 99 expert-generated metrics, ten of which over twenty-five percent of the expert interviewees agreed needed to be tracked. Triangulating this data, the historical data, and the new customer-generated top metric list resulted in a jump from 11% common metrics (historical vs. expert-generated) to 23% common metrics (customer-generated vs. expert-generated). Additionally, more of the 23% common metrics were common across multiple customers. The data also agreed with a rule of thumb written about by Brown (1996): "as a general rule, no individual employee should have to monitor more than 15 to 20 measures…a good number to shoot for is about a dozen, with half of those measures being the most important." Additionally Brown wrote, "The key to having a successful set of metrics is pairing down your database to the vital few key metrics that are linked to your success."

(2) How efficient and effective is commonalizing metrics in assessing performance?

The research did not provide a quantitative assessment to gage how much more effective or efficient an enterprise would be if metric commonality were adopted, however, there were qualitative benefits identified. Overall, commonalizing metrics across external stakeholders help drive program decisions in a more effective and more efficient way. Program decisions would be more effective because all customers would be tracking the same things, each with the same standard definitions, resulting in more of an "apples to apples" comparison of performance across all customers. Greater efficiency is achieved because the common data sharing results in less time and data interpretation to make the right program decisions. Experts and customers alike agreed with this assessment, and a majority of these two sets of stakeholders agreed that commonalizing metrics was a value-added tool to make decisions for this program.

(3) How do common metrics change over an operating life cycle?

This question was answered via the expert interviews. Of the 99 metrics generated by the experts, 53% of all metrics should be tracked regardless of the operating life cycle phase, while 47% of all metrics would provide more value if they were tracked at different life cycle phases. However, when investigating this same question for the "top ten" common expert-generated metrics, 90% of this subset should remain the same throughout the life cycle.

6.2: Future Research

This research has only touched started to address this new decision-making tool of metric commonality across an operating life cycle. There are four areas of research that should be explored further.

While it is important to "be lean," it is also important to not be too lean, and as such, further research could be explored in how much commonality is considered too much commonality, where the benefit of commonalizing metrics begins to adversely affect customer satisfaction (as they may view themselves as no longer being unique) or program decisions. Similar to commonality literature, perhaps a degree of commonality index could be created and applied to various enterprises that would be able to alert the enterprise of their "optimal amount of commonality."

The other three ideas draw upon the insights and reflections of Chapter 5. Leadership must be engaged if an enterprise is to adopt this new tool, and as such, it would be value-added to be able to quantify the benefits of metric commonality for their particular enterprise. Similarly, a more quantitative connection between metrics and strategic planning should be assessed as well. Finally, it is always important to take research concepts and apply them in other environments to show their adaptability. While this was inferred, it was not attempted.

Overall, while the results of this case study and this thesis may have only touched the "tip of the iceberg" in the concept of metric commonality, it is a good starting point in a new area of metric research and adds to the ever-expanding knowledge database of enterprise excellence.

REFERENCES

Alizon, Fabrice; Shooter, Steven B; and Simpson, Timothy W. "Henry Ford and the Model T: lessons for product platforming and mass customization." *Design Studies* 30.5 (September 2009): 588-605. Web.

Basili, Victor R; Caldiera, Gianluigi; and Romback, H. Dieter. "The Goal Question Metric Approach." *Encyclopedia of Software Engineering* 1 (1994): 528-532. Web.

Blackburn, Craig. "Metrics for Enterprise Transformation." *MIT Masters Thesis*. June 2009.

Boas, Ryan C. "Commonality in Complex Product Families: Implications of Divergence and Lifecycle Offsets." *MIT Doctor of Philosophy Thesis*. September 2008.

Bourne, Mike; Mills, John; Wilcox, Mark; Neely, Andy; and Platts, Ken. "Designing, implementing and updating performance measurement systems." *International Journal of Operations and Production Management* 20.7 (2000): 754-771. Web.

Brown, Mark Graham. *Keeping Score: Using the Right Metrics to drive World-Class Performance*. New York, NY: Productivity Press, 1996. Print.

Cote, V; Bourque, P; Oligny, S; and Rivard, N. "Software Metrics: An Overview Of Recent Results." *The Journal of Systems and Software* 8 (1988): 121-131. Web.

Dixon, J. Robb; Nanni, Jr., Alfred J; and Vollmann, Thomas E. *The New Performance Challenge: Measuring Operations for World-Class Competition*. Homewood, IL: Dow Jones-Irwin, 1990. Print.

Drucker, Peter F. "Managing for Business Effectiveness." *Harvard Business Review* (May-June 1963): 53-60. Web.

Eisenhardt, Kathleen M. "Building Theories from Case Study Research." *The Academy of Management Review* 14.4 (October 1989): 532-550. Web.

Fattah, Yuri and Stephens, Corey. *Commercial Aviation Safety Team (CAST) / International Civil Aviation Organization (ICAO) Common Taxonomy Team (CICTT) Overview/Outreach Briefing*. Web (www.intlaviationstandards.org).

Flapper, Simme Douwe P; Fortuin, Leonard; and Stoop, Paul PM. "Towards consistent performance management systems." *International Journal of Operations and Production Management* 16.7 (1996). Web.

Haddad, Marc G. "Knowledge Integration for Problem Solving in the Development of Complex Aerospace Systems." *MIT Doctoral Thesis*. May 2008. Pages 146-154.

Hammer, Michael. “The 7 Deadly Sins of Performance Measurement and How to Avoid Them.” *MIT Sloan Management Review* 48.3 (Spring 2007): 19-28. Web.

Hauser, John R. and Katz, Gerald M. “Metrics: You Are What You Measure!” *European Management Journal* 16.5 (October 1998): 517-528. Web.

Hubbard, Douglas K. *How to Measure Anything: Finding the Value of Intangibles in Business.* Hoboken, NJ: John Wiley & Sons, Inc., 2007. Print.

Hubbard, Douglas K. “The IT Measurement Inversion.” *CIO*, 13 June 2007. Web. 16 March 2010.

Ittner, Christopher D. and Larcker, David F. “Coming Up Short on Nonfinancial Performance Measurement.” *Harvard Business Review* (November 2003): 1-8. Web.

Jawahar, I.M. and McLaughlin, Gary L. “Toward a Descriptive Stakeholder Theory: An Organizational Life Cycle Approach.” *Academy of Management Review* 26.3 (2001): 397-414. Web.

Jaworski, Bernie and Kholi, Ajay. “Co-Creating the Voice of the Customer.” *Chapter 8 from the book titled The service-dominant logic of marketing: dialog, debate, and directions, by Robert F. Lusch and Stephen L. Vargo, editors.* Armon, NY: M.E. Sharpe, Inc. Pages 109-117. Web.

Jick, Todd D. “Mixing Qualitative and Quantitative Methods: Triangulation in Action.” *Administrative Science Quarterly* 24.4 (December 1979): 602-611. Web.

Keegan, Daniel P; Eiler, Robert G; and Jones, Charles R. “Are Your Performance Measures Obsolete?” *Management Accounting* 70.12 (June 1989): 45-50. Web.

Lean Advancement Initiative Website. http://lean.mit.edu.

Lohman, Clemens; Fortuin, Leonard; and Wouters, Marc. “Designing a performance measurement system: A case study.” *European Journal of Operational Research* 156 (2004): 267-286. Web.

Hubbard, Doulgas W. *How to Measure Anything: Finding the Value of Intangibles in Business.* Hoboken, NJ: John Wiley and Sons, 2007. Print.

Mahidhar, Vikram. “Designing the Lean Enterprise Performance Measurement System.” *MIT Masters Thesis.* September 2005.

Medori, David. “The development and implementation of an integrated performance measurement framework.” *Proceedings of Performance Measurement – Theory and Practice: International Conference, Cambridge, UK, 1998.* pp. 639-646.

Medori, David and Steeple, Derek. "A framework for auditing and enhancing performance measurement systems." *International Journal of Operations and Production Management* 20.5 (2000). Web.

Mingler, Paul. Personal discussion. 8 October 2009.

Murman, Earll, et al. *Lean Enterprise Value.* New York, NY: PALGRAVE, 2002. Print.

Neely, Andy; Richards, Huw; Mills, John; Platts, Ken; and Bourne, Mike. "Designing performance measures: a structured approach." *International Journal of Operations and Production Management* 17.11 (1997): 1131-1152. Web.

Neely, Andy; Mills, John; Platts, Ken; Gregory, Mike; and Richards, Huw. "Mapping measures and activities: a practical tool for assessing measurement systems." *Proceedings of the First International Conference of the European Operations Management Association, Cambridge, UK, 27-29 June 1994.* pp. 313-318.

Neely, Andy; Gregory, Mike; and Platts, Ken. "Performance measurement system design: A literature review and research agenda." *International Journal of Operations and Production Management* 15.4 (1995): 80-116. Web.

Nightingale, Deborah. *Metrics and Performance Measurement System for the Lean Enterprise.* Integrating the Lean Enterprise Class Presentation. Massachusetts Institute of Technology, Cambridge, MA. 19 October 2009.

Nightingale, Deborah and Rhodes, Donna. *Overview of Performance Management.* Enterprise Architecting Class Presentation. Massachusetts Institute of Technology, Cambridge, MA. 17 February 2010.

Nightingale, Deborah, et al. *Enterprise Strategic Analysis for Transformation.* MIT Lean Advancement Initiative, Cambridge, MA. September 2008.

Prasad, Biren. "Survey of life-cycle measures and metrics for concurrent product and process design." *Artificial Intelligence for Engineering Design, Analysis and Manufacturing* 14 (2000): 163-176. Web.

Schmenner, Roger W. and Vollmann, Thomas E. "Performance measures: Gaps, false alarms, and the 'usual suspects.'" *International Journal of Operations and Production Management* 14.12 (1994): 58-70. Web.

Sanderson, Susan and Uzumeri, Mustafa. "Managing product families: The case study of the Sony Walkman." *Research Policy* 24 (1995): 761-782. Web.

Simon, Herbert A. "Bounded Rationality and Organizational Learning." *Organization Science* 2.1 (1991): 125-134. Web.

Stephens, Corey; Ferrante, Olivier; Olsen, Kyle; and Sood, Vivek. "Standardizing International Taxonomies." *ISASI Forum* (January-March 2008): 8-13. Web.

Tuomela, Tero-Seppo. "The interplay of different levers of control: A case study of introducing a new performance measurement system." *Management Accounting Research* 16 (2005): 293-320. Web.

Valerdi, Ricardo and Blackburn, Craig. "Barriers to the Adoption of Measurement Systems." IEEE Working Paper, 2009. Personal hard copy.

Yin, Robert K. *Case Study Research Design and Methods, 3rd Edition*. Thousand Oaks, California: SAGE Publications, 2003. Print.

APPENDIX A:

Analysis of Strengths and Weaknesses of Performance Measurement Frameworks Mahidhar (2005)

Performance Measurement Framework	Strengths	Weaknesses
Strategic measurement and reporting technique (SMART)	• Integrates strategic objectives with operational performance measures. • Aggregates financial and non-financial measures across various functions and business units.	• Does not capture measures with respect to all stakeholder values • Does not provide any mechanism to identify causal relationships between measures across functions or levels. • Does not explicitly integrate the concept of continuous improvement. • May promote local optimization due to functional approach
The Balanced Score card	• Scorecard approach to integrate strategic, operational, and financial measures. • Focus on linkages and strategy maps • Most widely accepted	• The linkages between the measures are presumed and unidirectional. • Explicitly focuses on customers but leaves other stakeholders implicit. • No deployment system that breaks high-level goals down to the sub-process level .
European Foundation for Quality Management	• Contains self assessment tests • Focuses not only on the results, like the balanced scorecard, but also on the drivers of success	• Enterprise performance management is broader than quality management. • Loosely defined framework with no supporting process of implementation.
The Performance prism	• Has a much more comprehensive view of different stakeholders (e.g. investors, customers, employees, regulators and suppliers) than other frameworks. • Provides visual map causal relationship map of measures for individual stakeholders.	• It offers little about how the causal relationships between the performance measures are going to be realized. • There is little or no consideration is given to the existing systems that companies may have in place.
A Framework for design and audit	• Provides detailed implementation guidelines. It can be used both to design a new performance measurement system and to enhance an existing performance measurement system . • It also contains a unique description of how performance measures should be realized.	• The performance measurement grid provides basic design for the performance measurement system, and the grid is only constructed from six categories. • The causal relationships among the measures is not explained.
A Framework of factors affecting evolution	• Provides a systematic process of assessing the existing performance measurement system and adapting to the changing internal and external environment. • Design against people, process, system , technology	• Does not consider stakeholders as one of the factors affecting the measurement system.

APPENDIX B:
Expert Interview Questions

Objective:
Generate a list of expert-developed optimum performance metrics for the [product], spanning [product] lifecycle phases. Employees are to address the questions below; Alissa will take all responses and combine to generate one list of "recommended" metrics from the "experts."

Expert Typology:
What do you consider your area of expertise?

How comfortable are you in calling yourself an expert in this area?

How comfortable are you in calling yourself an expert in this area on [product]?

How many years have you been working (or worked) in your area of expertise?

How many years have you been working (or worked) in your area of expertise on the [product]?

Defining the Voice of the Customer:
Not elaborated upon so as to conceal identity of organization. There were two definitions of the voice of the customer (defined internally).

** NOTE: standard definitions would need to be established for the answers below, but not enough time is allotted in this interview period to review definitions **

Addressing the Voice of the Customer via [Product] Metrics through a Life Cycle:
What [product] performance measures do you believe are most effective in [addressing VOC #1] when the [product] is first entered into service?

What [product] performance measures do you believe are most effective in [addressing VOC #1] when the [product] is in the middle of its operating life (prime production)?

What [product] performance measures do you believe are most effective in [addressing VOC #1] when the [product] is preparing to be phased out?

What [product] performance measures do you believe are most effective in [addressing VOC #2] when the [product] is first entered into service?

What [product] performance measures do you believe are most effective in [addressing VOC #2] when the [product] is in the middle of its operating life (prime production)?

What [product] performance measures do you believe are most effective in [addressing VOC #2] when the [product] is preparing to be phased out?

Addressing Role of the Developer via [Product] Metrics through a Life Cycle:
What [product] performance measures do you believe are most effective in [helping you better do your job] when the [product] is first entered into service?

What [product] performance measures do you believe are most effective in [helping you better do your job] when the [product] is in the middle of its operating life (prime production)?

What [product] performance measures do you believe are most effective in [helping you better do your job] when the [product] is preparing to be phased out?

Effectiveness of Metric Commonality:
What do you believe is the optimal percentage of common [product] performance metrics across all customers that would result in maximum efficiency of understanding [product] performance?

What is your confidence interval of your answer above?

Motivating Factors for Commonalizing Metrics:
What data would you need to see to convince yourself that metric commonality is the right approach to managing [product] performance?

How much customer / developer interaction do you believe is necessary to determine these measures? What other stakeholders do you believe would need to be involved in this process?

What incentives or incentive structures should be in place to motivate the concept of metric commonality across customers?

Other:
I asked you what else you believe is important to the customer outside of [VOC #1 and VOC #2]. What [product] performance measures would you think should be in place to track this parameter?

Are there other programs you have worked on that you believe have a strong framework in determining proper [product] performance metrics? How successful do you believe those other programs are?

Do you have other comments or concerns you would like to discuss?

APPENDIX C:
Survey on Adoption of Metric Commonality Decision-Making Tool (adapted from Ricardo Valerdi)

Must-Be – referring to attributes where user is dissatisfied from its absence but never rises above neutral no matter how much of the attribute exists (i.e., good brakes).
One-Dimensional – referring to increasing user satisfaction from the presence of this attribute and decreasing satisfaction from its absence (i.e., gas mileage).
Attractive – indicates areas in which the user is more satisfied when the measurement system has the attribute but is not dissatisfied when it is absent; lack of an attribute leads to a neutral reaction (i.e., radio antenna that lowers into car body).

Please identify a single category for each attribute.

	Must-Be	One-Dimensional	Attractive
Well documented You are provided with documentation and training on how to adopt metric commonality principles for your enterprise.	☐	☐	☐
Trialability You can pilot the recording of the common metric set and, depending on its success, can implement this model as more of a standard.	☐	☐	☐
Low barrier of entry The transition between the way you record metrics now and the proposed way you should record metrics is not overly complex.	☐	☐	☐
Transparency There is easy access, as a product community, to this common metric data.	☐	☐	☐
Demonstrates value There is a clear link between this new model and its assumed value (higher performance and lower maintenance costs).	☐	☐	☐
Variety of Incentives The use of the new model includes personal incentives, or increases your job performance.	☐	☐	☐
Tailorable There is still opportunity for the metrics to be customized for your enterprise's particular needs.	☐	☐	☐
Information freshness The recorded data is updated at a predetermined periodic basis so that it continues to help drive decisions.	☐	☐	☐
Relative Advantage It has an advantage over the current metric recording process	☐	☐	☐
Compatibility It is compatible with your enterprise's current operating environment.	☐	☐	☐
On-going peer support It is a supported system (such as support provided through knowledgeable field representatives or IT experts if an IT-enabled system is developed).	☐	☐	☐
Credibility It is based on a method, approach, tool, or standard that has already proven itself to be valuable	☐	☐	☐

APPENDIX D:
Customer Interview Questions

Section 1:
What are your job's largest critical-to-qualities (CTQ's)?

How does tracking [product] performance integrate with your business objectives?

How did you create the list of [product] performance metrics you currently track?

What do you do with the [product] metrics once you record them?

Section 2:
What five to ten [product] performance metrics do you consider to be most important to address your job's CTQ's? Why?

Open discussion on data review (not listed in thesis)

Section 3:
From your standpoint, what are advantages and disadvantages to adopting metrics that other customers already use?

What sort of data would you want to see that would convince yourself that commonalizing metrics would benefit your [enterprise]?

What would improve, or incentivize, adoption of metric commonality?

Open discussion on adoption survey (Appendix C)

Are there other attributes we did not review that you believe should be considered when trying to adopt the model of metric commonality?