Navigating the Metrics Landscape: An Introductory Literature Guide to Metric Selection, Implementation, & Decision Making

Craig Blackburn and Ricardo Valerdi

Massachusetts Institute of Technology, Lean Advancement Initiative, United States, cdb@mit.edu & rvalerdi@mit.edu

Abstract

The focus of this paper is to depict the vast landscape of literature related to enterprise performance measurement in a concise and comprehensible manner for researchers and practitioners. We focus particularly on the enterprise as the unit of analysis and consider measurement systems from stakeholders at all levels. A broad range of considerations will be explored, ranging from micro-level considerations such as employee performance measurement to macro-level considerations such as enterprise measurement systems. Moreover, we discuss measurement-related problems identified in practice and solutions proposed in academic literature. To illustrate this evolution of measurement knowledge over time, we discuss the effects of metrics from three distinct viewpoints: (1) selecting the right metrics, (2) creating and implementing measurement frameworks; and (3) metrics for decision making.

Keywords – Metrics, Measurement, Performance Measurement, Decision Making

1 Introduction

The focus of this paper is to depict the vast landscape of literature related to enterprise performance measurement in a concise and comprehensible manner for researchers and practitioners. We focus particularly on the enterprise as the unit of analysis and consider measurement systems from stakeholders at all levels. To illustrate this evolution of measurement knowledge over time, we discuss the effects of metrics from three distinct viewpoints: (1) selecting the right metrics, (2) creating and implementing measurement frameworks; and (3) metrics for decision making.

First, we explore the idea of selecting the right metrics. In order to develop a common grounding, we expand on the concept of measurement and fundamental problems individuals and organizations face regarding measurement. This discussion focuses around common mistakes and metric selection methodologies, considering respective implications on individual behavior. We provide an example from professional baseball to demonstrate how thinking creatively can ensure metrics correspond to value added activities and increase human productivity.

Second, we describe the creation and implementation of measurement frameworks. Attributes of macro-level frameworks such as Kaplan and Norton's Balanced Scorecard will be compared with other complementary approaches [1]. We also discuss the many classifications of these frameworks, from "traditional" to "contemporary" systems, considering "structural" and "procedural" models, understanding temporal aspects, and identifying unique challenges and benefits from a case study of a micro (bottom-up) measurement system implementation.

Third, we discuss the role of metrics in decision making. In particular, we consider how to use metrics with imperfect information. To supplement various academic viewpoints provided, we offer a practical discussion regarding guidance for decision makers for focusing on the right problem and dealing with imperfect information – contextually relevant for managers.

This paper is an introductory guide for both practitioners and researchers to gain a better understanding about enterprise performance measurement. This guide is not intended to be collectively exhaustive, but indeed makes a point to articulate readings relevant to each section that one can consult for further information. In considering metrics selection, implementation and decision making there will never be a silver bullet – "a single development, in either technology or management technique, which by itself promises even one order of magnitude improvement in productivity, in reliability, in simplicity" [2]. The practical implications of all three metric subjects are highly dependent on a variety of factors, to include but not limited to: the maturity of an organization and their processes; top-down or bottom-up measurement system implementation; the industry being considered; the unit of analysis, such as people or business units; and the perspective taken during measurement. The principles and conclusions discussed in this paper will be depicted universally such that they can be applied in any context. For a brief overview of the literature discussed, Figure 1 provides a rough correlation regarding how select representative readings fit into each subtopic. This figure is relevant to how we discuss the work in the text of the paper, and should not be considered a complete classification of the work in question. Moreover, to supplement the literature breakdown, we provide metaphors and practical implication examples where appropriate.

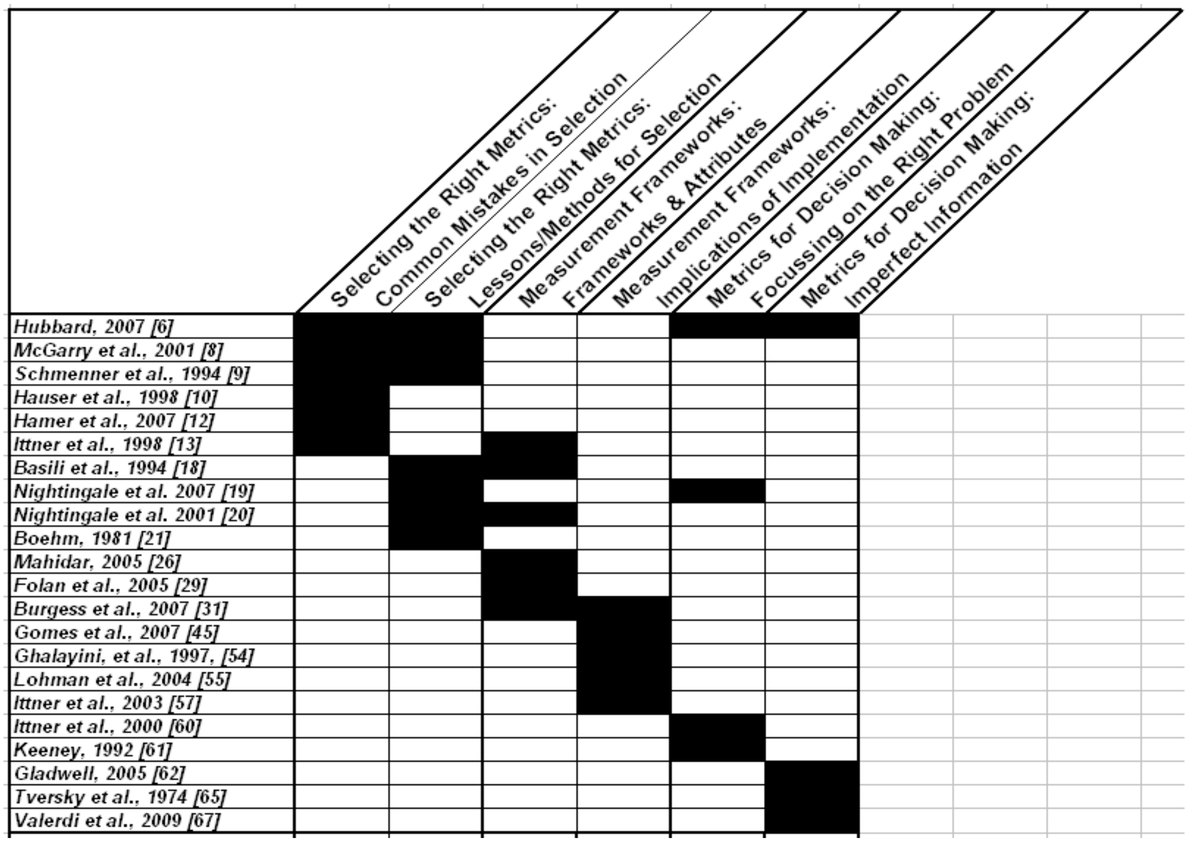

Figure 1 - *Literature Breakdown by Topic*

In the selection of these literary sources, works were chosen that emphasized macro-level measurement and we excluded works that focused on individual people or tasks. We drew from different disciplines including operations management, software engineering, aerospace engineering, product development, economics, accounting, etc. Within these disciplines we identified influential papers and books that provided fundamental ideas about performance measurement that could be generalized to other contexts.

2 The Importance of Metrics

A great wealth of research has been performed on selecting the right metrics and behaviors that metrics will encourage. The importance of metrics is a topic that has been studied for over half of a century, dating at least as far back as 1956, when *Administrative Science Quarterly* published two articles about concerns with dysfunctional responses to performance measurement [3; from 4; and 5]. Currently, research on this topic has spread throughout many industries and a mature body of knowledge has been developed. This section consolidates much of this research and identifies some critical lessons that can be generalized for any industry of interest. First, we need to define what exactly a metric is and the problem space of interest. According to Hubbard (2007), a metric can be defined as an *observation that reduces uncertainty wherein the result is expressed as a quantity [6].* Thus, traditional manufacturing metrics such as cycle time and defect density are considered as well as more abstract concepts, such as culture and employee involvement. Similarly yet more formally, Kitterman (2005) defines a metric as a *quantified value of an attribute, obtained by the process of assigning numerical values to attributes, which is compared to what is expected [7].* Furthermore, McGarry et al., (2001) considers a measure *the objective data collected to fulfil the information needs of the manager* [8]. The similarity between these definitions is that we are quantifying attributes in support of decisions.

There are many popular heuristics that authors employ to concisely articulate fundamental key principles of metric selection. For example: Schmenner and Vollmann (1993) identify the old adage "What gets measured, gets managed;" [9] the well known Kaplan and Norton (1992) Balanced Scorecard claim's that "What you measure is what you get;" [1] and Hauser and Katz (1998) state that "You are what you Measure" [10]. The general point is that these heuristics have repeatedly shown that metrics drive behavior in people. Ultimately, these simple heuristics are combined with lessons from previous literature to provide: (i) common metric selection mistakes; (ii) methods for metric selection; and (iii) how metrics relate to value identification.

2.1 Common Metric Selection Mistakes

Picking the wrong metric is easy. In the most rudimentary manner of expressing this measurement problem, many reward certain actions while hoping an unrelated and often contradictors results [11]. For instance, in professional sports if a player has incentives built into his contract where he gets bonuses based on the amounts of points he scores alone, that player is thus encouraged to be selfish and hence will diminish the potential of the overall team as a system. The same is true for performance metrics in education – wherein awarded grades influence employment, higher learning, tuition reimbursement and parental respect – yet those who award grades are trying to fulfil the goal of knowledge transfer from teacher to student [11]. Similarly, professors are expected to pursue excellence in teaching yet are rewarded on publications [11]. In addition to not considering the consequences of a metric on human or system behavior, a collection of prominent performance metric mistakes has been articulated in an array of literature, in a non-exhaustive list identified below.

1. *Not using the right measure (ignoring something important) or choosing metrics that are wrong (i.e. for a phone help service, customers don't just want quick answers, they want accurate ones as well) [8; 9; and 10]*
2. *Having metrics reflect functions as opposed to cross-functional processes [12; and 13]*
3. *Assuming one knows what is important to measure without giving enough thought or using measures that intentionally make you look good [12]*
4. *Measuring only a part of what matters, measuring from your view rather than the customers, or forgetting your goal [10; and 12]*
5. *Implementing metrics that focus on short-term results, or that do not give thought to consequences on human behavior and enterprise performance [8; 9; 10; 11; 12; and 13].*
6. *Having metrics that are not actionable or hard for a team/group to impact or collecting too much data [8; 10; and 13]*

Making any of these critical mistakes could counteract any good intention of standing-up a metric. From inspection of these common mistakes, three simple themes emerge: (1) measuring correct and complete value added activities, not just easy to measure attributes; (2) considering the effects of standing-up a metric on individual and team behavior – how they will respond to the metric in the absence of management intervention; and (3) fostering a culture of commitment to measurement and cross-company collaboration. This breakdown is deconstructed as follows (numbers corresponding to the measurement mistakes list).

- Value Added
  - [1] – Ignoring Something Important
  - [4] – Measuring only part of what matters
- Behavioral Effects
  - [5] – Not considering effect on humans
  - [6] – Hard for a team/group to impact
- Commitment
  - [2] – Company boundaries dictate metrics
  - [3] – Not being serious about measurement

The implications of falling victim to these common metric selection mistakes can be quite profound. First, not measuring correct and complete value added measures effectively can lead to sub-optimization. It is important not to measure because you can, measure because it helps inform decision making about activities that add value to your company, your suppliers, and your customers. As an example, one can easily think too narrow and only measure part of what matters. In baseball, if a team measures only power maximization because it helps offense and fans enjoy homeruns, production may suffer as players become less agile and cannot steal bases or run as well. Even though teams may not openly emphasize it as much, they also need players that can run the bases fast and play great defense. Thus, when considering signing a player all these attributes need to be taken into account, not just power or easily measurable factors. Similarly, there are many interdependencies that cannot be ignored in all professions, thus one needs to ensure metrics are aimed measuring correct and complete value added activities.

Second, with respect to implementing metrics that actually cause the individual and behavioral effect desired, consider the example from professional sports. Understand that if your rewards are risky (pay for individual as opposed to team performance), counter-productive behavior will become the norm. Conversely, if you attempt to reward a team on a factor that is hard to control – like if a team rewards a player only based on cumulative team performance – the player will not be motivated to succeed since they will understand their individual efforts have a negligible causal relationship to their potential for reward. Thus, the problem lies in determining optimal metrics that that link an individuals' contribution to team success. Baseball's statistician guru, Bill James, sought to quell this dilemma with the creation of "win shares" – a complex metric derived from both traditional and non-traditional metrics that together contribute to portraying a players' contribution to team performance [14].

Third, not being committed or serious about measurement can lead to disaster. If one uses metrics designed to make people look good, real problems will not be exposed until it is too late to address them. It would be naïve for managers to assume they know with absolute certainty what is right to measure, as they should constantly seek feedback from their regarding how metrics are affecting both behavior and enterprise performance. Furthermore, if traditional organizational boundaries dictate performance metrics and these factors are never audited, the true value of the selected metrics will be compromised. For example, baseball scouts traditionally measured talent by raw speed, power, and gut feeling [15]. A significant cultural movement was needed for industry analysts to think outside the box to reconsider metrics and value, an example that will be revisited in section 2.3.

2.2 Metric Selection Methodologies

Aside from avoiding these common pitfalls and mistakes, other broad criteria for metric selection has been proposed as well. Nightingale (2007) offers that metrics need to be: strategic, to align behavior with company objectives; quantitative, to provide understanding of progress towards these objectives; and qualitative, to provide organizational understanding as to how the metric is valuable [16]. Implicit in these criteria is also the need for metrics to be actionable, thus a performance measurement system needs to help depict what needs to be done and by whom. Further, performance measurement systems should show that one is doing the right job (meeting stakeholder requirements) and doing the job right (being economically resourceful) [16].

Implementing these broad criteria and recommendations is often easier said than done, as often portraying a metric's value to employees or measuring intangible factors presents hurdles. In addition to the identification of the common measurement mistakes and criteria for effective metrics, research has been performed to identify some of the most over-measured and under-measured phenomenon in organizations. Schemmenner & Vollmann (1994) performed a study across senior manufacturing executives to analyze which of the twelve items of greatest interest to them were over-measured or under-measured [9]. Ninety-two executives rated how important they thought each measure was for long-term success and then articulated the degree to which current performance measures were either inhibiting or supporting them as summarized in Table 1.

Table 1 - *Commonly Over-Measured & Under-Measured Metrics Relative to Importance on Long-Term Performance [9]*

Under-Measured	Equal	Over-Measured
Employee Involvement	*Integration with Customers*	*Machine Efficiency*
Customer Satisfaction	*Overhead Cost Reduction*	*Labor Efficiency*
New Product Introduction	*Volume Flexibility*	*Direct Cost Reduction*
-	*Throughput Times*	-
-	*Quality*	-
-	*Computer Systems*	-

As one may have guessed, the most under-measured factors for long-term success are "softer" or more intangible metrics. Conversely, the most over-measured factors are those that are, albeit arguably, easier to measure. This important assumption leads to one main question – why are we not measuring these factors? Hubbard (2007) provides some insight to this question [6]. First, he recalls from his experiences that often "costs are measured more than the more uncertain benefits" and "small 'hard' benefits are measured more than large 'soft' benefits." Moreover, the author identifies some common rationales given for under-measurement (or not measuring), including: measurement being too expensive, or perhaps that the resulting statistics would not be useful enough to support decision making. To these objections, four basic assumptions influence the measurement of "softer" factors: (1) your problem is not as unique as you think;" (2) you have more data than you think; (3) you need less data than you think; and (4) there is a useful measurement that is much simpler than you think. Considering the factors themselves, he adds "if it matters at all it is detectable and observable," "if it is detectable, it can be detected as an amount," and "if it can be detected as a range of possible amounts, it can be measured" [6]. Taking this advice, we venture into some common methodologies to assist with the problem of metric selection.

Given the common mistakes and intangible nature of some critical under-measured long-term success factors, the next question is: how does one choose the most appropriate metric? Some of the general guidelines proposed for metrics selection relate to the common mistakes discussed above. Although many researchers and practitioners converge on the appropriate steps for metric selection, different delivery styles exist. A sample of previously identified metric selection methodologies can be seen in Appendix A. As a generalization extracted from inspection of Appendix A, we submit that there are four critical steps for metric selection, seen in Figure 2, and expanded upon below.

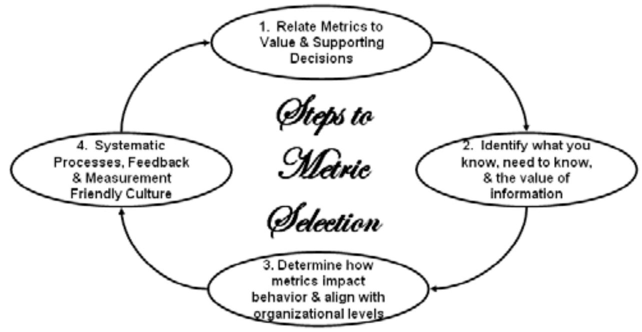

Figure 2 - *The Steps to Metric Selection*

Step 1: Identify what you are trying to measure, the decision it is trying to support, and how it is part of a greater purpose. This is the first step in identifying the right metric is one which helps avoid the aforementioned mistake of not relating metrics to value added activities. In this step, one needs to identify the stakeholder, their needs and consider how the metric will support the decision making process. Moreover, one needs to identify what decision the metric will support and how strongly it relates to a fundamental core goal, such as customer satisfaction. Some systems and software engineering measurement communities have endorsed Vic Basili's Goal-Question-Metric (GQM) approach to identify metrics, wherein you: (1) *identify the information goal* stakeholders want to know and why, working top-down, including organizational and project goals; (2) *ask the question* that will aid in evaluating if the goal is being met; (3) *determine the measures* that need to be gathered to collect information to answer the question; and (4) *apply* the selected metrics and *evaluate* their usefulness [8; and 19; from 18]. Similarly, the Lean Advancement Initiative's Lean Enterprise Self Assessment

Test (LESAT) X-Matrix tool compliments and could add more structure to using this method, as it requires stakeholders to formally document links between metrics, strategic objectives, stakeholder values, and key processes [20].

Step 2: Determine what you know, and what you need to know - the value of unknown information. Now that you have defined the selected metric, one needs to determine how much information they actually know that can reduce uncertainty. Then, consider how much information needs to be acquired to optimize the decision making process relative to the cost of gathering it. Hubbard (2007) provides a step in his methodology that accounts for this need: determine the value of information – the consequences of being wrong and the chance of being wrong, evaluate what degree of measurement effort is justifiable [6]. Another perspective regarding value of information decision guidelines comes from Boehm (1981), wherein five conditions are identified under which it would make sense to investigating alternatives prior to committing to a course of action [21].

1. There are alternatives whose payoff varies greatly, depending on some critical states of nature.
2. The critical states of nature have an appreciable probability of occurring.
3. The investigations have a high probability of accurately identifying the occurrence of the critical states of nature.
4. The required cost and schedule of the investigation do not overly curtail their net value.
5. There exists significant side benefits derived from performing the investigations.

Boehm (2002) adds that one method of comparing risk exposure (RE) mitigation alternatives is by using risk reduction leverage (RRL) techniques [22]. Using this method of comparison, alternatives are evaluated quantitatively by taking the difference between risk exposures before and after pursuing an alternative, and dividing this term by the implementation cost of the alternative, seen in Figure 3. Thus, the higher the risk reduction leverage value of an alternative, the more attractive it is since risk is reduced most efficiently relative to cost.

$$\text{RRL} = \frac{\text{RE}_{\text{BEFORE}} - \text{RE}_{\text{AFTER}}}{\text{RISK REDUCTION COST}}$$

Figure 3 – *Risk Reduction Leverage Equation [22]*

Step 3: Understand who is impacted by the metric, and how it aligns vertically with different levels of the organization. This step is associated to the mistake of not considering how metric selection will affect individual or group behavior, but is more encompassing on an enterprise level. For example, not only does one consider the impact of the metric on individuals and groups, but it is also necessary to ensure the company's values are aligned to its customers' and suppliers' values. Moreover, metrics that are selected should be aligned vertically in the organization as well, relating incentives for factory floor level workers to core company goals. This topic of alignment will be expanded on in section 3.1.

Step 4: Have a systematic process and a measurement friendly culture; evaluate timeliness of information, quality, and whether or not the system has working feedback. This final step relates to the issue of not being serious about measurement and provides insight into implementation. In addition to collecting and disseminating metric information in a timely fashion, there needs to be an active feedback loop to ensure that when change occurs it is depicted in the metric. One method for method for considering this step would be with a thermostat approach, wherein periodic feedback informs management which metrics to emphasize based on the enterprise's current performance and where historical data dictates they need to be to ensure short-term and long-term success [23]. When this approach is applied correctly, managers and employees can constantly be focussing on only a few metrics, those which are easiest to manipulate to improve profitability. With a culture that is cognizant about the value of a metric and proactively seeking improvement, one can optimize their chances of success at metric selection.

2.3 Identifying Value with Metrics

Metric selection is important in many aspects, from one's personal life to their work life. Despite the varying contextual nature of metric selection, proper execution of the first step – identifying what you are trying to measure and ensuring it related to value added – is most paramount for success. Blackburn and Valerdi (2008) provide an example from professional baseball that articulates the need to define value appropriately, understanding that traditional industry-accepted metrics are not always the most appropriate [24].

The example provided stems from Michael Lewis' book *Moneyball* (2004) – a story depicting one of the most over achieving baseball teams of all time, the Oakland Athletics, and their methods for success [15]. These methods stem from the value system implemented by their General Manager, Billy Beane. In professional baseball, measures of performance historically were based on what was easiest to measure, essentially the individual's sole contribution to run production in the form of runs, hits, homeruns, and batting average. Beane viewed baseball's performance metrics much differently, as he saw the offensive potential of the team as dependent on an integrated system wherein everyone had to do their part to manufacture runs – in an assembly line fashion. Starting from scratch, Beane's first step to metric selection was to understand what fundamental metric could embody the maximization of team scoring potential. Given that an offense in baseball can be limited only by outs (events), rather than time – Beane emphasized the need for his offense to *not get out,* and thus he was one of the first managers to use more holistic performance metrics that were in-line with the eventual goal of team victories (predicted by walks, on base percentage, slugging percentage, not getting caught out on the base paths, etc..).

Even though this example may seem somewhat unusual, it reinforces the need for one to first explore the problem space before moving to the solution and understanding value before determining the metric. Concluding, one critical lesson learned from this example is that if traditional organizational boundaries and mechanisms do not facilitate value identification, you can't be afraid to go against the grain and be the Billy Beane of your organization.

3 Measurement Frameworks

Although value identification and the steps to effective metric selection provide valuable insight, many have taken these established principles one step further and created measurement frameworks – more detailed models for guiding metric collection through implementation in the decision making process. Fundamentally, a measurement framework can serve two purposes within an organization, to measure and to motivate [25]. Considering these factors, most frameworks are as concerned with aligning efforts and influencing behaviors (motivating) as they are about determining the as-is state and the trajectory of the organization (measuring or monitoring). In addition to these roles, Mahidar (2005) adds that a performance measurement system serves three additional needs, all five listed below [26].

- Monitor: measure & record actual performance.
- Control: to identify & close the gap between target & actual performance.
- Improvement: to identify improvement opportunities.
- Coordination: determining information decision makers need (leading indicators) & facilitating both internal communication (across processes) as well as external communication (amongst stakeholders).
- Motivate: Align behavior & encourage transformation.

Different from measurement selection, measurement frameworks have their own unique and contextual attributes that need to be considered for implementation, as there is no one size fits all solution. Comparable to literature regarding metric selection, the field of identification and discussion of metric frameworks is highly oversaturated. Many classical macro-level approaches to performance measurement use an interrelationship of performance criteria, such as the seven depicted in the Sink and Tuttle (1989) model [28; from 27].

- Effectiveness: right job at the right time.
- Efficiency: resources consumed.
- Quality: throughout the enterprise perspective.
- Productivity: ratio of output to input.
- Work life Quality: needed for performing systems.
- Innovation: to sustain and improve performance.
- Profitability/budgetability: the ultimate goal.

Other methods used for deriving performance measurement frameworks involve breaking top business process into groups – such as (i) primary processes, (ii) support processes, and (iii) development processes – and then nesting detailed metrics within each category [28]. Another common derivation of this performance measurement categorization style has been to segregate economic considerations from the rest of the system – perhaps by using (i) economic factors, (ii) external relations, (iii) internal relations, and (iv) ability to change – as primary corporate analysis units [28]. Common issues practitioners often incur when evaluating their performance measurement system pertain to either configuring a typology for performance measurement management similar to those above, or vertically linking performance metrics to value and corporate strategy.

Considering these common issues, many disciplines draft their own professional guidelines to forge communities with specialized contextual knowledge. For example, to synthesize the most relevant concepts in the field of systems engineering, the INCOSE Measurement Working Group prepared the Systems Engineering Measurement Primer (1998) [19]. This primer serves as an introduction to the process of how to consider metrics from the more focused lens of the systems engineering practitioner. Thus, in this section we provide a brief look into: commonly used cross-disciplinary measurement frameworks and their respective attributes; and implications of implementing a bottom-up measurement framework derived from a case study of an aerospace defense contractor, highlighting both significant benefits and challenges.

3.1 Measurement Frameworks & Attributes

Performance measurement frameworks can help an organization determine measurement boundaries, direct measurement dimensions or views, and provide insight regarding relationships amongst the performance measures themselves [30; from 29]. Burgess et al. (2007) performed an extensive literature review within the field of performance measurement discussing the difference between "traditional" and "contemporary" performance measurement systems – wherein the eventual shift was generated by the desire of many to move away from older financial-based measures in favor for more balanced systems that incorporate an array of non-financial metrics [31]. Complimentary research suggests that popular high-level financial metrics such as return on investment capital can be used to compare a company's overall strategy to net income, but concedes that this metric (and similar financial ones) is often inappropriately applied to evaluate project performance or employee appraisal [32].

Similarly, other research supports this suggestion and notes that aggregate financial-based metrics (such as economic value added) fell out of favor with many operational decision makers for a variety reasons, notably: they are too complex, they provide little or incomplete information on key drivers of future performance, they make it difficult to consider softer metrics like human or intellectual capital, and they often do not correspond to shareholder return as hoped [13]. Generally, human capital and softer metrics are associated with higher performance when an enterprise's strategy is market differentiation based, the product line is complex, the environment is uncertain, the industry is knowledge based, or core human capital is scarce [33]. Since most industries are faced with some combination of these factors, we submit that non-financial performance measures will be relevant in most measurement contexts.

Noting this objection to financial measurement and the existence of others, we acknowledge that non-financial measurement systems are not without flaw. Many opponents of non-financial measures believe that financial metrics facilitate trade-offs in decision making most objectively, and fear that measurement systems will begin to rely on flawed subjective metrics, and have other objections as well [13].

In addition to the shift from financial to non-financial measurement focus, other analytical work identified the evolution of two types of performance measurement frameworks – (1) structural frameworks, which specify a typology for performance measure management, and (2) procedural frameworks, which are step-by-step processes for developing performance measures from strategy [29]. These types of frameworks when combined form a more complete performance measurement system. The authors conclude from their studies that the development of structural frameworks is maturing faster than procedural frameworks. Helping close the maturity gap between structural and procedural frameworks, we believe an extensive amount of research and implementation has been performed this past decade, for example regarding Six Sigma's DMAIC (Define, Measure, Analyze, Implement, and Control) procedural measurement framework [34]. We also suggest that the steps to metric selection identified above can be used as an example of a procedural framework.

Furthermore, some commonly used frameworks referenced thus far in this paper include: the Practical Software Measurement (PSM) approach's Measurement Construct that provides a structural framework for managing high-level performance measures [8]; and the GQM approach that embodies a procedural approach by directly linking measures through strategy in just three steps [18]. Mahidar (2005) identifies four additional structural frameworks and two procedural frameworks, which we describe through their respective strengths and weaknesses in Attachment B for reference [26]. A breakdown of these frameworks from Attachment B with some others discussed in this paper is seen in Table 2.

Table 2 - *Performance Measurement Framework Typology*

Structural	Procedural	Both
Strategic Measurement & Reporting Technique [35]	A Framework for Design & Audit [39]	The Balanced Scorecard [1]
The Performance Prism [36]	A Framework for Factors Affecting Evolution [40]	Extended Enterprise Balanced Scorecard (Structural) and Procedural Frameworks [29]
European Foundation for Quality Management – EFQM [37]	Define-Measure-Analyze-Implement-Control [34]	-
PSM's Measurement Contstruct [8]	GQM [18]	-
Value Stream Mapping [38]	Steps to Metric Selection	-

Considering the frameworks above and performance measurement needs articulated through literature, we submit that there are five common attributes to a complete performance measurement system.

1. Alignment of metrics, both (i) vertically from corporate vision to operational execution and (ii) horizontally to consider the stakeholder satisfaction (suppliers, customers, community, etc…).
2. Improvement of internal processes.
3. Innovation, learning, and growth.
4. Feedback from all levels of the organization.
5. Temporal tense – depicting historical performance, the present state, and predicted future direction.

To supplement the first four attributes more directly empirically derived from the listed frameworks, the fifth was inspired from the most pressing need identified from other literature. Brown (1996) identified the need for three temporal perspectives to be conveyed through performance measurement systems – portraying the historical, current, and future performance of the company [41]. Dixon (1990) shares a similar sentiment, noting that performance measures need to be dynamic to keep pace with the ever changing business environment [42]. Melynk, et al., (2004) furthers this common theme by asserting that metrics have a temporal "tense" [43]. The two tenses the authors identified are: (1) outcome-oriented (lagging) indicators from which by analyzing the past one can improve the future (most financial metrics); and (2) predictive (leading) indicators that are likely to correlate strongly to outcomes of interest. Some of the non-financial measures considered in the frameworks discussed above can be used as meaningful leading indicators, such as customer satisfaction.

Although not explicitly referenced in the five attributes, financial concerns should be embedded into frameworks one develops to provide the right balance. The identified attributes should be viewed as general driving factors to consider when developing or using a performance measurement framework. Given these critical performance measurement framework attributes, we offer some simple universal methods and tools for practical implementation regarding attributes one through five.

With respect to vertical alignment, the GQM method for aligning corporate strategy to all levels of the organizational hierarchy would be an effective method of alignment. For horizontal alignment, lean principles provide methodologies and frameworks for determining critical stakeholders and value stream mapping to coordinate organizational movement to that of customers, suppliers, and other stakeholders [38]. The Lean Advancement Initiative's aforementioned X-Matrix tool could also assist with both horizontal and vertical alignment processes, as it facilitates linking metrics to strategic objectives, stakeholder values, and key processes – identifying gaps in the process where metrics could be implemented [20]. In addition to assisting with value stream mapping and alignment, lean principles and the proactive pursuit of continuous improvement could be used for improving internal processes. With respect to innovation, companies in all business environments need to pursue value opportunities – "the lure of greater success via market and/or technology opportunities prompts transformation initiatives" – before being in a position of a value crisis – "steadily declining market performance, cash flow problems, etc., prompt recognition that transformation is necessary to survive" [44]. A commitment to innovation could be demonstrated in a variety of ways, perhaps by fostering a culture of employee innovation and engagement, or by investing in research and development. Lastly, the existence of a feedback loop gives the performance measurement system evolutionary life and promotes continuous improvement (as referenced in section 2.2). Over time, iterations of stakeholders input regarding how the system enables or inhibits operations can be considered in parallel with the information needs of senior leadership to improve system utility.

3.2 Implications of Implementing Bottom-Up & Top-Down Measurement Frameworks

Evolving from "traditional" to "contemporary" performance measurement systems, the integration of "structural" and "procedural" models, and considering the temporal aspect of measurement frameworks – Gomes, et al., (2007) provide an in depth review of performance measurement literature, proclaiming the emergence of two distinct implementation themes: (1) the "universal" theme, which includes approaches to performance measurement and implementation methodologies which advocate transferability across organizational context and operating environments; and (2) the "contingency" theme, which includes approaches which stress the unique characteristics of organizations, functions, and/or business units in relation to performance measurement and implementation methodologies [45]. From the generalized conversations of metric selection and performance measurement, we have gravitated towards the "universal" concept, as we suggest that the themes and principles discussed are transferable across multiple contexts. The "contingency" approach has focused on implementation issues, focussing on three dimensions: (i) individual performance measures; (ii) performance measurement systems; and (iii) the relationship between the performance measurement system and the environment in which it operates [45; from 46]. As we have already provided insight on the first two dimensions, we briefly discuss the relationship between the system itself and a specific contextual environment – the implementation of measurement frameworks, recommendations, challenges and benefits.

In the practical context of a large company, measurement systems are traditionally implemented from two contrasting perspectives: (1) a macro (top-down) perspective by executives striving to align the company to a corporate strategy; or (2) from a micro (bottom-up) perspective by mid-level managers trying to manage the day to day operations of programs, matrix organizational function departments, or perhaps a factory site. In designing a measurement system for implementation (micro or macro), Burgess et al. (2007) identifyies the following critical characteristics for success [31].

- linking to the business strategy [42; and 47];
- linking measures hierarchically from strategy through to operational detail [42; and 48];
- balanced measures such as financial and non-financial [49] and internal and external [50];
- the system should be easy to understand, be simple to use and provide timely information [42; and 48];
- providing a feedback mechanism to enable the corrective actions and flow of information to decision-making function of the company [51];
- allowing ongoing updating and changes as needed [52].

As these ideas are consistent with other principles unveiled in this paper, we will not go further but direct you to Burgess, et al., (2007) for more information [31]. Macro framework implementation also coincides with the universal principles that have been discussed thus far, so we refrain from reiterating the general challenges and attributes of measurement systems.

Considering micro-level framework implementation, Ghalayini, et al., (1997) introduce the integrated dynamic performance measurement framework that integrates: (1) management; (2) process improvement team; and (3) the factory floor operations [31; from 53]. For contextual understanding of this interaction, Blackburn and Valerdi (2009) provide a case study of an aerospace defense contractors' enterprise measurement system to gain a better understanding of this integration mechanism [54]. To summarize, the micro-level framework described in Blackburn and Valerdi (2009) provides a comparable function to that described by Ghalayini et al., (1997). Table 3 below provides insight into some of the opportunities and challenges regarding this integration process, as identified in the study.

Table 3 - *Performance Measurement System Opportunities & Challenges [54]*

Opportunities	Challenges
Fresh & plentiful data for decision makers	Employee engagement: considering the aging workforce and a lack of wide-spread enthusiasm for learning a new system
Visibility into all value stream and program activities	Resistance to measurement: wherein individuals are suspicious as to how the data may be being used to assign blame
Constant communication amongst stakeholders	Information Misinterpretation: the second order effects of management having an extensive information gathering system – leading management to sometimes prematurely draw conclusions that may not necessarily be accurate
Being able to hold people accountable & provide feedback where appropriate	Information saturation: the information overload that can occur for managers and touch laborers, which results in stakeholders forgetting which metrics are most important and value added due to the plethora of information they need to process
Fostering a culture of continuous improvement	-
Employee engagement & empowerment (considering the potential of positive momentum)	-

As performance measurement trends grow to include more of an extended enterprise, including the supplier base and other stakeholders, problems such as these will become increasingly highlighted. Additioanlly, other challenges for growing extended enterprises pertain to:

- decentralized reporting structures;
- deficient insight in cohesion between measures;
- uncertainty about what to measure;
- poor communication between reporters and users; and
- dispersed information technology infrastructure [56; from 55].

In short, as one considers what measurement framework they desire to adapt to their organization, or how to design their own performance measurement system, it is most imperative that they consider: (1) lessons learned from previous implementations; (2) the opportunities and challenges of measurement frameworks; (3) any necessary adaptation to the organization to implement a system that is consistent with their goals and working environment – be it structural or procedural, traditional or contemporary, past or present, or even macro or micro. Finally, we provide an example of the potential impact one might be able to achieve with a performance measurement system.

In one study, researchers hypothesized and that the implementation of a measurement system would lead to performance improvement. To validate this theory, financial successes of companies that built causal models linking metrics to strategy were examined compared to those who did not (a total of 157 companies) [57]. As expected, the companies that used such models (23%) did significantly better with respect to both return on assets (+2.95%) and return on equity (+5.14%) than their counterparts.

We warn that although this study infers that implementing performance measurement systems corresponds to significantly better financial performance than non-using competitors, it is highly plausible that the previously higher performing firms would have been more likely to adopt these contemporary measurement methods [58]. Thus, the need to explore similar and more focused studies regarding whether this relationship can be attributed to causation or covariation would be needed for definitive conclusions. However, regardless of the strength of causation, the positive correlation indicates that more benefit than harm could come from implementing performance measurement systems.

4 Metrics for Decision Making

Metrics and measurement frameworks have no meaning if they are not used to make decisions. The practical reality is that managers have to make decisions with imperfect information. In light of this problem, modern management techniques are beginning to emphasize "management by means" and the evaluation of relationships and processes as opposed to traditional "management by results" that focuses on the outcomes of processes [59]. Johnson and Broms (2000) warn that traditional quantitative thinking limits the perception of the decision maker to one dimension, despite nature having many alternative dimensions oft forgotten. The authors argue that traditional quantitative analysis stems too much from the study of mechanistic systems with definitive properties, whereas organizations are living entities with interactions and relationships that traditional methods cannot quantify. Thus, denouncing traditional mechanistic quantification techniques, managers are faced with the difficult task of determining what modern quantification methods work best and identifying the role of metrics in decision making. Considering this quandary, in this final section we expand on two final themes for decision makers: tying decisions to the right problem and being confident in making decisions with minimal or imperfect information.

4.1 Focusing on the Right Problem.

In the metrics selection discussion, we presented the idea of mapping metrics to part of a greater ideal. Specifically, one needs to ensure that metrics correlate to the decisions they are supporting. In practice, three methods are commonly used for understanding value drivers: (1) intuition; (2) standard classifications – financial, internal business processes, customer, learning and growth; and (3) statistical analysis of leading and lagging indicators of financial performance, which can allow decision makers to identify statistically supported weights for the most important metrics with regards to how they relate to financial performance [60].

Before using one of these methods; however, decision makers need to step back, explore the problem space, and ensure that they are focussing on the right problem regardless of the method chosen. Thus, Hubbard (2007) proposes questions decision makers should consider before measuring to help them focus on the right problem [6], and Nightingale (2007) identifies questions for assessing a performance measurement system to ensure that it is focussing on the right problem as well [16].

- *Questions to consider before measuring [6]:*
  1. What is the decision this [measurement] 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?
- *Questions to asses a measurement system [16]:*
  6. Is the right information received at the right time?
  7. Are the metrics tied to the organization's goals?
  8. Does it identify root causes?
  9. Does it consider all stakeholders and their needs?
  10. Does it motivate individual or group action as intended?
  11. Does it accurately portray progress?
  12. Is it easy to use?

The first set of questions focus on exploring the problem space, ensuring decision makers consciously consider if there is a causal nature between the measure and desired action or if there is just an association confounded by other factors. Once the problem space has been explored and one has verified that their metrics are focussing on the right problem, the second set of questions should assist in ensuring decision makers are gathering the right information for effective decision making (representative of the right problem, actionable, timely, etc…). We further suggest that one consider the granularity of detail necessary to support their decisions, as well as the cost of false positives (acting/intervening when you believe there is a problem but there is not a problem) or false negatives (not acting/intervening when there actually is a problem).

Too often, decision makers jump to solutions without understanding the causal factors – which leads to either the aforementioned false positives or negatives. Keeney (1992) considers this dilemma, first differentiating between what he identifies as "alternative-focused thinking" and "value-focused thinking" [61]. According to Keeney:

Value-focused thinking involves starting at the best and working to make it a reality. Alternative-focused thinking is starting with what is readily available and taking the best of the lot.

With this thought, we assert that too often the focus is on easy-to-measure 'hard' data rather than 'soft' objectives like goodwill, quality of the product, amount to be learned, or societal benefit. Keeney identifies the sequence of decision making events in alternative-focused thinking as follows: (1) recognizing a decision problem, (2) identifying alternatives, (3) specifying values, (4) evaluating alternatives, and (5) selecting an alternative. As a generalization, in value-focused thinking, specifying values would occur before alternatives are selected. This improvement in thinking will help one avoid rushing to conclusions and hence false positives or negatives.

4.2 Decision Making with Imperfect Information

Now that we have given ample consideration to the need to explore the problem space, we discuss the phenomenon of acting on imperfect or negligible information. Previously, we discussed how the drawback of information overload could perhaps jeopardize one's ability to make a meaningful decision. Conversely, Gladwell (2005) conveys that often the best decisions are made by relying on a few pieces of high quality information, rather than endless databases [62]. This theme is depicted through a war exercise, wherein an experienced Marine Officer acting as a Middle Eastern rogue combat group was able to continually outsmart his adversaries acting on behalf of the United States. The Marine Officer was able to use a few meaningful pieces of information through the "fog" of war to outsmart his adversaries who had a wealth of data, perhaps too much, to act on. Consistent with this example, some suggest that the acquisition of new data or information for the decision maker can easily lead to a more uncertain or hazardous state [63]. By moving away from the critical pieces of information traditionally relied upon in favor of databases of information, the potential to make a decision that helps or hurts the cause is expanded [63]. As it will be harder for decision makers to interpret the meaning of a larger group of metrics or some ambiguous aggregate number, the potential to focus on the wrong metrics or see a relationship that doesn't exist increases.

In order to make snap judgements with the success of the Marine Officer and avoid the decision maker's dilemma when too much information is available, research suggests that a major component of decision making lies with knowledge appraisal – the extent to which one can determine data quality [64]. In knowledge appraisal, one needs to determine the extent to which available information describes the context of concern, and evaluate if biases have been eliminated to the greatest controllable extent. Another context in which experts are often pressed to make decisions involving incomplete information and knowledge appraisal is in project management. Estimation forecasts are desired early in a project lifecycle regarding critical attributes, such as cost, schedule, and effort. Thus, managers are often left with few choices – parametric predictive models or expert opinion.

In particular, we are interested in the nature of the expert opinion since there is more ambiguity and need for direction. Using expert knowledge as the fundamental measurement strength, we need to account for potential weaknesses – such as cognitive biases. In particular, some well studied biases include: (i) anchoring, wherein one relies on specific information or an arbitrary value and influences judgment; (ii) the "halo/horns" effect, wherein if people fist see one attribute that predisposes them to favor or disfavor one alternative, they are more likely to interpret subsequent information in a way that supports their conclusion; (iii) bandwagon bias, wherein the presence of others and their interpretations affects one's judgment; (iv) hindsight bias, where people exaggerate what could have been anticipated in foresight and consider the outcome having been relatively inevitable; and (v) overconfidence or optimism bias, where people exaggerate the confidence and completeness of their own knowledge [6; 64; and 65]. Although the first four biases can be systematically reduced when one is cognizant of them, we propose that elimination of optimism bias proposes a more difficult task. In light of this problem, general calibration techniques of experts and corresponding exercises to combat this effect have been studied [6; and 66]. Valerdi and Blackburn (2009) applied these techniques on systems engineers to understand the affects of optimism bias on the profession and provide insight regarding optimism reduction [67]. Insight from this study confirms that systems engineers are as susceptible to optimism bias as any other profession, and that the effects can be quelled when they are made cognizant of this bias through calibration exercised and given advice regarding methods of bias mitigation. Hence, when managers in any field are relying on expert judgement to make decisions, simple awareness of the problem can lead to more accurate knowledge appraisal.

In short, when acting with imperfect information, decision makers need to be dependent on their experts or just a few pieces of critical information to help them act appropriately. Being cognizant of the human element of decision making, including biases, will help one interpret information in the most effective manner possible.

5 Conclusions

In this paper, we have discussed the basics of metric selection, performance measurement frameworks, and the role of metrics in decision making. Practical examples of how one can choose good or bad metrics were given, and we discussed some of the most commonly over-measured and under-measured attributes associated with long-term industry success. To think about value, an example from professional baseball was provided that showed how one can use non-traditional methods to link metrics to value. Next, we discussed the many classifications of performance measurement frameworks, from "traditional" to "contemporary" systems, considering "structural" and "procedural" models, understanding the temporal aspects, and even identifying unique challenges and benefits of a micro (bottom-up) implementation as opposed to a macro (top-down) framework. Finally, we discussed how to think about metrics in the decision making process. This final section emphasized that we need to employ an open-minded value-focused approach to understanding value before engaging in measurement and discussed how decisions can be made with a marginal amount of information.

We note that although the majority of topics discussed in this paper have been extensively researched, they have not been synthesized in an organized form. This is partly due to the broad range of domains that write about metrics but fail to identify common issues in other domains. As a result of this disconnect between disciplines, the same mistakes are being repeated. We hope this paper can enable further integration of lessons learned from operations management, software engineering, aerospace engineering, product development, economics and accounting with respect to measurement.

In conclusion, we have provided a central source one can use to find resources that will further assist them with exploring their own metrics challenges. Within this exploration, we personally hope this collection has provided a better understanding of the oversaturated field of metrics literature – to give you a map, and help you make some sense of your own metrics journey.

6 References

[1] Kaplan, R. and Norton, D., The Balanced Scorecard - Measures that Drive Performance, *Harvard Business Review*, Vol. 70, Iss. 1, pp. 71-80, 1992.

[2] Brooks, F. P., *The Mythical Man Month*, New York, NY: Addison Wesley, 1995.

[3] Kelman, S., and Friedman, J., "Performance Improvement and Performance Dysfunction: An Empirical Examination of Impacts of the Emergency Room Wait-Time Target in the English National Health Service," Working Paper Series rwp07-034, Harvard University, John F. Kennedy School of Government, 2007.

[4] Berliner, J. S., "A problem in soviet business management," *Administrative Science Quarterly* Vol. 1, Iss. 1, pp. 86–101, 1956.

[5] Ridgeway, V. F., "Dysfunctional consequences of performance measurements," *Administrative Science Quarterly*, Vol. 1, Iss. 2, pp. 240–247, 1956.

[6] Hubbard, D. W., *How to Measure Anything: Finding the Value of "Intangibles" in Business,* Hoboken, NJ, Wiley, 2007.

[7] Kitterman D., "A Structured Method for Generating, Evaluating and Using Metrics," *15th Annual International Symposium of INCOSE*, Rochester, NY, July 2005.

[8] McGarry, J., Card, D., Jones, C., Layman, B., Clark, E., Dean, J. and Hall, F., *Practical Software Measurement: Objective Information for Decision Makers*, Addison-Wesley, 2001.

[9] Schmenner, R. W., and Vollmann, T. E., "Performance Measures: Gaps, False Alarms and the "Usual Suspects,"

International Journal of Operations and Production Management, Vol. 14, No. 12, pp. 58-69, University Press, 1994.

[10] Hauser, J. R. and Katz, G., Metrics**:** you are what you measure!, *European Management Journal*, Vol. 16, No. 5, pp. 517-528, October 1998.

[11] Kerr, S., "On the folly of rewarding A, while hoping for B," *Academy of Management Executive*, Vol. 9, No. 1, pp. 7-14, 1995.

[12] Hammer, M., Haney, C. J., Wester, A., Ciccone, R. and Gaffney, P., The 7 Deadly Sins of Performance Measurement and How to Avoid Them, *MIT Sloan Management Review*, Vol. 48, Iss., 3, Spring 2007.

[13] Ittner, C. D. and Larcker, D. F., "Innovations in Performance Measurement: Trends and Research Implications," *Journal of Management Accounting Research*, Vol. 10, pp. 205-238, 1998.

[14] James, B. and Henzler, J., *Win Shares*, Morton Grove, IL, STATS Inc., 2002.

[15] Lewis, M., *Moneyball: The Art of Winning an Unfair Game*, New York, NY, W.W. Norton & Co., 2004.

[16] Nightingale, D., "Metrics and Performance Measurement System for the Lean Enterprise," Integrating the Lean Enterprise, Massachusetts Institute of Technology, 17 October 2007.

[17] Harbour, *The Basics of Performance Measurement,* New York, New York, Productivity Press, 1997.

[18] Basili, V., Caldeira, G. and Rombach, H. D., "The Goal Question Metric Approach," in J. Marciniak (ed.), *Encyclopedia of Software Engineering*, Wiley, 1994.

[19] Systems Engineering Measurement Primer: A Basic Introduction to Systems Engineering Measurement Concepts and Use, Version 1.0, International Council on Systems Engineering (INCOSE), March 1998.

[20] Nightingale, D., Cool, C., Brown, K., Mize, J., Shields, T. and Hallam, C. *Lean Enterprise Self-Assessment Tool Facilitator's Guide (v1.0)*. Cambridge: Massachusetts Institute of Technology, Lean Advancement Initiative, 2001.

[21] Boehm, B. W., *Software Engineering Economics*, Upper Saddle River, NJ, Prentice Hall, 1981.

[22] Boehm, B. W., "Software Risk Management: Overview and Recent Developments," Tutorial, COCOMO/SCM Forum #17, Los Angeles, CA, 22 October 2002.

[23] Hauser, J. R., Metrics Thermostat, *The Journal of Product Innovation Management*, Vol. 18, Iss. 3, pp. 134-153, 2001.

[24] Blackburn, C. and Valerdi, R., "Measuring Systems Engineering Success: Insights from Baseball," *18th Annual International Symposium of INCOSE,* the Netherlands, June 2008.

[25] Mintzberg, H. *The Structuring of Organizations,* Englewood Cliffs, N.J, Prentice-Hall, 1979.

[26] Mahidar, V., "Designing the Lean Enterprise Performance Measurement System," *MIT Masters Thesis*, September 2005.

[27] Sink, S. and Tuttle, T., "Planning and Measurement in your Organization of the Future," *Industrial Engineering and Management Press*, Norcross, GA, 1989.

[28] Rolstadas, A., "Enterprise Performance Measurement," *International Journal of Operations & Production Management*, Vol. 18, No. 9/10, pp. 989-999, 1998.

[29] Folan, P. and Browne, J., "Development of an Extended Enterprise Performance Measurement System," *Production Planning and Control*, Vol. 16, No. 6, pp. 531-544, September 2005.

[30] Rouse, P. and Putterill, M., "An Integral Framework for Performance Measurement," *Management Decision*, Vol. 41, No. 8, pp. 791-805, 2003.

[31] Burgess, T. F., Ong, T. S., and Shaw, N. E., "The Prevalence of Performance Measurement System Types," *International Journal of Productivity and Performance Management,* Vol. 56, No. 7, pp. 583-602, 2007.

[32] Kessler, B., Patterson, R., Roth, G. and Nightingale, N., "A White Paper on Using ROIC as the LAI Top Level Metric," Lean Advancement Initiative, Massachusetts Institute of Technology, 2003.

[33] Huselid, M. A. and Barnes, J. E., Human Capital Measurement Systems as a Source of Competitive Advantage, Working Paper, School of Management and Labor Relations Department of Human Resource Management, Rutgers University, 16 April, 2003.

[34] De Feo, J. A. and Barnard, W., *JURAN Institute's Six Sigma Breakthrough and Beyond - Quality Performance Breakthrough Methods*, New York, NY, Tata McGraw-Hill Publishing Company Limited, 2005.

[35] Cross, K. F. and Lynch, R., "The SMART way to Define and Sustain Success," *National Productivity Review*, Vol. 8, No. 1, pp. 23-34, 1988.

[36] Neely, A., and Adams, C., "The Performance Prism Perspective," *The Journal of Cost Management*, Vol. 15, No. 1, pp. 7-15, 2001.

[37] Jackson, S., "Exploring the Suitability of the European Foundation for Quality Management (EFQM) Excellence Model as a Framework for Delivering Clinical Governance in the UK National Health Service," *Quality Assurance Journal,* Vol. 5, 19-31, 2001.

[38] Murman, E., Allen, T., Bozdogan, K., Cutcher-Gershenfeld, J., McManus, H., Nightingale, D., Rebentisch, E., Shields, T., Stahl, F., Walton, M., Warmkessel, J., Weiss, S., and Widnall, S., *Lean Enterprise Value: Insights from MIT's Lean Advancement Initiative,* Basingstoke, Palgrave, 2002.

[39] Medori, D., and Steeple, "A Framework for Auditing and Enhancing Performance Measurement Systems," *International Journal of Operations & Production Management*, Vol. 20, Iss. 5, pp. 520-533, 2000.

[40] Kennerly, M., and Neely A., "Measuring Performance in a Changing Business Environment," *International Journal of Operations & Production Management*, Vol. 23, Iss. 2, pp. 213-229, 2003.

[41] Brown, M. G., *Keeping Score: Using the Right Metrics to Drive World Class Performance*, New York, NY, Productivity Press, 1996.

[42] Dixon, J., Nanni, A. and Vollmann, T., *The New Performance Challenge: Measuring Operations for World-Class Competition*, Dow Jones Irwin, Homewood, IL, 1990.

[43] Melnyk, S., Stewart, D. M., and Swink, M., "Metrics and performance measurement in operations management: dealing with the metrics maze," *Journal of Operations Management*, Vol. 22, No. 3, pp. 209-217, 2004.

[44] Rouse, W. B., "A Theory of Enterprise Transformation," *Systems Engineering*, Vol. 8, No. 4, pp. 279-295, 2005.

[45] Gomes, C. F., Yasin, M. M. and Lisboa, J. V., "An Empirical Investigation of Manufacturing Performance Measures Utilization," *International Journal of Productivity and Performance Management*, Vol. 56 No. 3, pp. 187-204, 2007.

[46] Neely, A., Gregory, M. and Platts, K., "Performance measurement system design: a literature review and research agenda", *International Journal of Operations & Production Management*, Vol. 15 No. 4, pp. 80-116, 1995.

[47] Keegan, D .P., Eiler, R. G. and Jones, C. R., "Are your performance measures obsolete?", *Management Accounting*, Vol. 70, No. 12, pp. 45-50, June 1989.

[48] Lynch, R. L. and Cross, K. F., *Measure up – The Essential Guide to Measuring Business Performance*, Mandarin, London, 1991.

[49] Feurer, R. and Chaharbaghi, K., "Performance measurement in strategic change", *Benchmarking for Quality Management and Technology,* Vol. 2 No. 2, pp. 64-83, 1995.

[50] Waggoner, D., Neely, A. and Kennerley, M., "The forces that shape organizational performance measurement systems: an interdisciplinary review", *International Journal of Production Economics*, Vol. 60/61, pp. 53-60, 1999.

[51] Bititci, S., Carrie, A. S. and McDevitt, L., "Integrated performance measurement system: a development guide", *International Journal of Operations & Production Management*, Vol. 17 No. 5, pp. 522-34, 1997.

[52] Ghalayini, A. M. and Noble, J. S., "The changing basis of performance measurement," *International Journal of Operations & Production Management*, Vol. 16 No. 8, pp. 63-80, 1996.

[53] Ghalayini, A. M., Noble, J. S. and Crowe, T. J., "An integrated dynamic performance measurement system for improving manufacturing competitiveness", *International Journal of Production Economics*, Vol. 48 No. 3, pp. 207-25, 1997.

[54] Blackburn, C., and Valerdi, R., "Practical Implementation of an Enterprise Measurement System: From Inception to Transformation," *7th Annual Conference on Systems Engineering Research*, Loughborough, United Kingdom, April 2009.

[55] Lohman, C., Fortuin, L. and Wouters, M., "Designing a Performance Measurement System: A Case Study," European Journal of Operational Research, Vol. 156, Is. 2, pp. 267–286, 2004.

[56] Folan, P., Higgins, P. and Browne, J., "A Communications Framework for Extended Enterprise Performance Measurement," *International Journal of Computer Integrated Manufacturing*, Vol. 19, No. 4, pp. 301-314, June 2006.

[57] Ittner, C. D. and Larcker, D. F., "Coming up Short on Nonfinancial Performance Measurement," *Harvard Business Review*, pp. 88-95, November, 2003.

[58] York, K. M. and Miree, C. E., "Causation or covariation: an empirical link between TQM and financial performance," *Journal of Operations Management*, Vol. 22, pp. 291-311, 2004.

[59] Johnson, H. T. and Broms, A, *Profit Beyond Measure*: *Extraordinary Results Through Attention to Work and People*, Free Press, New York, NY, 2000.

[60] Ittner, C. D. and Larcker, D. F., "Non-financial Performance Measures: What Works and What Doesn't," *Financial Times*, Sec. Mastering Management, 16 October, 2000.

[61] Keeney, R. L., *Value-Focused Thinking: A Path to Creative Decisionmaking*, Cambridge, MA, Harvard University Press, 1992.

[62] Gladwell, M., *Blink: The Power of Thinking Without Thinking*, Back Bay Books, New York, NY, 2005.

[63] Sisson, J. C., Schoomaker, E. B. and Ross, J. C., "Clinical Decision Analysys: The Hazard of Using Additional Data," in Arkes, H. R. and Hammond, K. R. (Eds), *Judgment and Decision Making: An Interdisciplinary Reader,* (pp. 354–363), Cambridge, England, Cambridge University Press, 1982.

[64] Fischhoff, B., "Debiasing," in Kahneman, D., Slovic, P. and Tversky, A. (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 422–444), Cambridge, England, Cambridge University Press, 1982.

[65] Tversky, A., Kahneman, D., Judgment under Uncertainty: Heuristics and Biases, *Science*, Vol. 185, No. 4157, pp. 1124-1131, 1974.

[66] Lichtenstein, S., Fischhoff, B. and Phillips, L. D., "Calibration of Probabilities: The State of the Art to 1980," in Kahneman, D., Slovic, P. and Tversky, A. (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 306–334), Cambridge, England, Cambridge University Press, 1982.

[67] Valerdi, R., and Blackburn, C., "The Human Element of Decision Making in Systems Engineers: A Focus on Optimism," *19th Annual International Symposium of INCOSE,* Singapore, July 2009.

7 Biography

Craig Blackburn is a Research Assistant in the Lean Advancement Initiative. He is also a Masters student in the Department of Aeronautics and Astronautics, and in the Engineering Systems Division at the Massachusetts Institute of Technology. He received his B.S in Mechanical Engineering from The Pennsylvania State University in 2007 and is a 2nd Lieutenant in the United States Air Force. His current research interests include metrics for lean enterprise transformation.

Ricardo Valerdi is a Research Associate at the Lean Advancement Initiative at Massachusetts Institute of Technology and a founding member of the Systems Engineering Advancement Research Initiative. He is also a Visiting Associate at the Center for Systems and Software Engineering at University of Southern California (USC). He earned his BS in Electrical Engineering from the University of San Diego, MS and PhD in Industrial & Systems Engineering from USC. Dr. Valerdi is a member of INCOSE and serves on its Board of Directors. He is also co-Editor-in-Chief of the Journal of Enterprise Transformation,

8 Attachments

8.1 Attachment A: Sampled Metrics Selection Methodologies

	Steps to Creating Effective Metrics				
	Hubbard [6]	Huser & Katz [10]	Hammer [12]	Harbour [17]	Kitterman [7]
1	What are you trying to measure? What is the real meaning of intangible?	Listen to the customer, who is it and what do they need, what outcomes are metrics trying to improve	Decide what to measure	Accurately measure key performance variables (productivity, quality, timeliness, cycle time, resource utilization, and costs)	What decision are you trying to support?
2	Why do you care - what is the decision you are trying to support	Understand the job - what do managers and employees value, how do decisions effect metrics and outcomes?	Decide how to measure - precision: carefully and well defined (units, range, etc...)	Include a comparative basis to assist in understanding the performance level	Is it possible to take action and what actions are necessary
3	How much do you know - what ranges or probabilities represent your uncertainty	Understand the relationships - internal customers/suppliers, tradeoffs. Make sure everyone is considered	Decide how to measure - accuracy: any metric is a fraction of an ideal (customer satisfaction), close gap between metric and reality	Metrics need to be collected and distributed on a timely basis	Would the action create a change?
4	What is the value of information? Consequences/chance of being wrong? Justify measurement effort	Understand the linkages. Align actions and decisions with long-term company goals	Decide how to measure - overhead: inexpensive and simple works	Metrics need to be analyzable on both a macro and micro basis	If the change occurs, will it show up in a future value of the metric
5	What observations can confirm/eliminate possibilities?	Test Correlations and Manager/Employee Reaction	Decide how to measure - robsutness: Design around manipulation and unintended behaviors	Metrics cannot be manipulated to achieve desired results	-
6	Account for avoidable errors, value of information	Involve Managers and employees - gather input and feedback, do not use information for attacks	Use metrics systematically - embed in a disciplined process	Ensure measres are SMART - Specific, Measurable, Action-Oriented, and Timely	-
7	-	Seek new paradigms - Focus on Ouputs not Inputs. Look at policy and understand the enterprise perspective.	Create a measure friendly culture - don't use for infighting or blame. Personal role modeling, reward, implementation, commitment, articulation	-	-

8.2 Attachment B: Structural & Procedural Measurement Frameworks [26]

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.