An Abridged Enterprise Assessment Model to Promote Consistent Reassessment:
Model Development, Assessment Process and Results Analysis

by

L. Nathan Perkins

Bachelor of Art in Neuroscience,
University of Southern California, May 2007

Submitted to the Engineering Systems Division
in Partial Fulfillment of the Requirements for the Degree of

Master of Science in Technology and Policy

at the
Massachusetts Institute of Technology
June 2011

Signature of Author: _____
L. Nathan Perkins
Engineering Systems Division
11 May 2011

Certified by: _____
Ricardo Valerdi
Research Associate, Engineering Systems Division
Thesis Supervisor

Certified by: _____
Deborah J. Nightingale
Professor of Aeronautics and Astronautics and Engineering Systems
Director, Center for Technology, Policy and Industrial Development
Thesis Reader

Accepted by: _____
Dava Newman
Professor of Aeronautics and Astronautics and Engineering Systems
Director, Technology and Policy Program

This page is intentionally left blank.

An Abridged Enterprise Assessment Model to Promote Consistent Reassessment: Model Development, Assessment Process and Results Analysis

by

L. Nathan Perkins

Submitted to the Engineering Systems Division
on May 11, 2011
in Partial Fulfillment of the Requirements for the Degree of
Master of Science in Technology and Policy

Abstract

Enterprise assessment is increasingly important, both as a cross-time and cross-industry measurement and as a guiding force in enterprise transformation. Assessments provide crucial information about strengths, areas for improvement and potential investment strategies for achieving performance benefits. As performance is being recognized as a complex and multifaceted construct, assessment tools seek to incorporate and reflect a holistic measurement of performance across multiple dimensions such as stakeholder value, leadership, culture and quality.

The Lean Enterprise Self-Assessment Tool (LESAT) is one such enterprise assessment tool that closely ties into a clearly defined enterprise transformation framework and roadmap. Ideal use of assessment involves regular reassessment of 54 practices and continual feedback, but due to the resource and time commitment required to perform assessment, this iterative process is deprioritized. In order to facilitate and promote regular reassessment, we demonstrate a methodology for creating an abridged assessment tool. By creating a predictive model based on the unidemnsionality of LESAT, a small selection of highly indicative practices is used to predict the remaining practices. Based on these predictions, respondents assess follow-up practices selected to target high-priority areas for improvement.

Using this approach, we are able to create an abridged LESAT that assesses six of the original 54 practices for the predictive model and an additional twelve dynamically selected practices to target high-priority areas. Based on training data and novel testing data (271 respondents from 24 companies), we validate the accuracy of the predictive model and show that high-priority areas are correctly identified over 90% of the time. The abridged LESAT shows promise as a way to reassess, with significantly lower time and resource commitment normally required.

We review the practical applications of the abridged LESAT and present a revised recommended process for assessment and for evaluation of results. The revised process seeks to articulate how the new assessment tool can be practically applied in the context of an ongoing enterprise transformation.

Thesis Supervisor: Dr. Ricardo Valerdi
Title: Research Associate, Engineering Systems Division

This page is intentionally left blank.

Acknowledgments

"Nihil est... simul et inventum et perfectum." – Cicero

I would like to thank a number of people who have contributed, both directly and indirectly, to my work and to the writing of this thesis. This thesis is possible because of the guidance, feedback and insights contributed by those around me.

First, I am truly grateful to both Dr. Ricardo Valerdi and Professor Deborah Nightingale who have provided support and guidance over the course of my graduate education. They have contributed their friendship and their professional knowledge, contributing both to my academic work and my personal growth. I am forever indebted.

I am also grateful for many other colleagues throughout the Lean Advancement Initiative, who have patiently listened to my ideas and provided valuable feedback. LAI is host to many great minds, and I feel lucky to have been able to work among them. Thanks Leyla, JK, Jorge, Tom, John and Jordan, to name just a few.

On a personal note, I am forever grateful to my parents who have continually provided encouragement and support through all my adventures. They have been an inspiration. Thanks Penny and Bill! Likewise, I am appreciative of the support of my close friends, who were patient with me as I toiled away, and provided encouragement when needed.

This page intentionally left blank.

Table of Contents

List of Tables

This page is intentionally left blank.

List of Figures

This page is intentionally left blank.

Chapter 1. Introduction

In an era of rapid innovation and a changing global landscape, enterprises and organizations must continually redefine their mission, value and competencies to achieve a competitive advantage (Rouse, 2005). Local optimization and process improvement have long been crucial to business performance (Hoerl, 1998; Siviy, Penn, & Stoddard, 2007), but with fewer competitive boundaries and global competition, enterprise leaders must now look beyond local optimization in order to achieve competitive differentiation and maintain market share (Denison & Mishra, 1995). Enterprise leadership increasingly look beyond traditional performance metrics such as bottom-line costs and profit margins (Bourne, Franco, & Wilkes, 2003), incorporating factors such as culture, quality, stakeholder value, leadership commitment, process flow and organizational learning into a multidimensional measure of performance (Maskell, 1991).

By approaching success in from a holistic, multidimensional vantage point, enterprises are able to differentiate themselves from competition, increase stakeholder satisfaction, and achieve long-term financial success. To realize this new paradigm of success and to achieve the requisite cultural shift, enterprises must transform their values, processes and strategy (Denison & Mishra, 1995). These transformation efforts that seek to redefine the corporate culture and achieve increased value realization are often unsuccessful (Kotnour, 2011). Failed transformation stems from a variety of factors, including leadership commitment, resource allocation, lack of vision and/or lack of enterprise knowledge (Kotter, 1995; Rouse, 2011).

To overcome barriers to transformation roadmaps, new tools have been developed. Transformation roadmaps provide guidance and clearly defined steps to clarify transformation

goals, to ensure adequate planning and resource allocation, and to align the requisite knowledge and capabilities (Nightingale, 2009). The development of transformation roadmaps was complemented by the development of enterprise assessment tools, which allowed leaders and transformation planners to measure and evaluate enterprise performance on range of criteria. Enterprise assessment tools help expose objective information regarding current performance and enterprise maturity, as well as help guide the transformation planning process by identifying key weaknesses or areas for improvement (Nightingale & Mize, 2002). These tools enable enterprise to better achieve transformation, when they were adopted and tightly integrated with strategic planning and resource allocation processes (Hallam, 2003).

Despite the value and success of assessment tools, case studies suggest that such tools are often underused or simply used only once due to the associated time and resource commitment (Abdimomunova, Perkins, Valerdi, Shields, & Nightingale, 2011). Cost-conscious enterprise leaders forgo repeat assessment or full adoption of assessment tools because of the time and effort that goes into the assessment process. Yet, by not having detailed information on enterprise maturity, capabilities and weaknesses, the transformation process is impaired by a lack of enterprise knowledge (Perkins, Abdimomunova, Valerdi, Shields, & Nightingale, 2011). Many organizations lose out on potential performance improvements and opportunities for competitive differentiation due to shortcuts made in the assessment and transformation planning process.

1.1. Research Goal & Hypotheses

Motivated by the underutilization of enterprise assessment tools, this thesis aims to create an assessment tool and process that provides actionable enterprise knowledge without the substantial time and resource commitment required by ongoing assessments. As assessment tools

have many inputs reflecting a range of principles and best practices, the aim is to infer information about enterprise performance from a smaller set of indicative practices. Based on the inferences made, potentially high-priority or actionable principles and practices would be identified and manually assessed to confirm the accuracy of the inferences. This two step process (assessing indicative practices, then confirming potentially high-priority practices) will be referred to as an abridged assessment, as it has the potential to provide crucial inputs into the transformation planning process without the burden of assessing all principles and practices in the full assessment.

Subsequent chapters will present a methodology for building an abridged assessment tool from a full assessment, and applies this methodology to an existing enterprise assessment tool. This thesis seeks to confirm three hypotheses regarding the existence and efficacy of abridged assessment: (1) an abridged assessment can use a predictive model to identify, target and confirm high-priority areas for improvement; (2) the standard assessment process can incorporate the abridged assessment enabling enterprises to increase assessment frequency without proportionally increasing the resource burden; and (3) the abridged assessment can yield key insights related both to improving performance and potential areas for further improvement.

1.2. Research Methodology

To investigate the three hypotheses proposed, this thesis applies a novel technique for generating an abridged assessment tool and demonstrates the technique using an existing enterprise assessment tool. In order to disprove the first null hypotheses, we use a quantitative research methodology (Creswell, 2002). By assuming that assessment respondents would provide consistent scores to practices irrespective of the order and number of practices in the assessment tool, we can retrospectively simulate an abridged assessment (the treatment) and

directly compare the results of the abridged assessment to the full data from the same respondent (the control). By altering the order and practices presented to each respondent based on more limited inputs, the abridged assessment seeks to identify a subgroup of high-priority practices that aligns with what the full assessment would have also identified as high-priority.

The second and third hypotheses are addressed through literature review of existing documentation on the assessment process and assessment insights (Dellinger & Leech, 2007). By integrating existing literature on the assessment process and assessment analysis, we hope to identify potential conflicts and/or changes needed to accommodate the abridged assessment tool.

1.3. Thesis Organization

To lay the groundwork, we will briefly review the organization and progression of the thesis.

The second chapter provides context and scope for the thesis by defining enterprise assessment and providing an overview of the range and variety of enterprise assessment tools that exist.

The third chapter focuses in on the Lean Enterprise Self-Assessment Tool (LESAT), a specific enterprise assessment tool. In this chapter, we introduce the tool, review literature on the recommended assessment process and discuss insights that can be gained from the assessment process. This additional background detail on LESAT is provided as a basis for subsequent chapters in which we develop and validate an abridged version of LESAT.

The fourth chapter provides a detailed description of the methodology for creating an abridged assessment. The methodology proposed in generalizable, and can be applied to a range of enterprise assessment tools that have quantitative or numeric inputs.

Having established a general methodology for creating abridged assessment, we apply the methodology to LESAT in the fifth chapter. The chapter describes what practices are included in an abridged version of LESAT, and uses retrospective statistical analysis to validate the ability of the abridged LESAT to identify relevant information for enterprise transformation.

The sixth chapter discusses the implications and applications of an abridged LESAT by returning to the recommended process and potential insights introduced in the third chapter. We reconcile the new abridged assessment tool with the process and analysis strategies that are recommended for the full assessment. This chapter provides guidance on how the abridged version of LESAT could be practically deployed as part of an enterprises strategic planning cycle. In addition to discussing the practice, we also touch on the limitations and threats to the abridged LESAT. By incorporating limitations of the abridged assessment into to the process, it is possible to make the recommended process more resilient to the limitations abridged assessment tool.

Finally, in the seventh chapter we return to the hypotheses to determine whether the abridged assessments are viable tools that can be used in an enterprise context to identify high-priority areas for improvement.

Chapter 2. Enterprise Assessment

Before introducing the abridged assessment methodology, it is important to have sufficient background information on enterprise assessment and its increasing importance, both as a tool for cross-time and cross-industry measurement and as a guiding force in enterprise transformation. Assessments provide crucial information about strengths, areas for improvement and potential investment strategies for achieving performance benefits. As performance is being recognized as a complex and multifaceted construct, assessment tools seek to incorporate and reflect a holistic measurement of performance across multiple dimensions such as stakeholder value, leadership, culture and quality.

With a growing range of tools available, this chapter introduces four prevalent assessment strategies as examples of different approaches to enterprise assessment. The tools are compared in terms of use cases, principles measured, outputs and contextual factors. Due to a lack of causal evidence between the principles assessed and objective evidence of improved performance/transformation outcomes, evaluating an assessment tool based solely on effectiveness is almost impossible. Instead, finding a contextually relevant tool that can be integrated into the strategic process should drive an enterprise's choice of assessment tool. By committing to a single tool and establishing closed-loop feedback structures (Hallam, 2003), enterprises are able to achieve the internal knowledge and historical data necessary to improve performance and drive ongoing transformation efforts. These enterprise assessment tools and the idea of closed-loop feedback are important concepts in establishing an effective abridged assessment that aids in guiding the transformation effort.

2.1. Key Terms: Enterprise and Assessment

Before addressing the concept of enterprise assessment, it is important to simply clarify what is meant by the terms *enterprise* and *assessment*. Having clearly established definitions as a starting place will help ground the ongoing discussion of the importance of assessment within the context of enterprises.

First, an *enterprise* is "a complex, integrated and interdependent system of people, processes and technology that creates value as determined by its key stakeholders. An enterprise typically consists of multiple organizations (e.g., suppliers, partners, regulators) rather than a single corporation, division or government unit" (Nightingale & Srinivasan, 2011). Unlike organizations, the concept of an enterprise provides a holistic, end-to-end vantage point for understanding value creation. It provides a paradigm for free from traditional structural boundaries in order to gain more complete understanding of organizational roles, stakeholders and value creation.

Second, *assessment* refers to a comprehensive set of tools and processes for both measuring and evaluating a subject – for our purposes, an enterprise – in order to guide performance (Roth, 1999). In this way, assessment seeks to establish quantitative or qualitative measurement tools linked with the desired outcomes. These measurements then act as inputs for an evaluation process, integrating the measurement inputs with contextual factors and desired outcomes to appraise performance. The measurement phase acts to reinforce the desired principles and ground the evaluation process in terms of real data (Hauser & Katz, 1998), while the evaluation phase acts to subjectively interpret the performance data in a broader context and translate that interpretation into specific areas/topics for future improvement.

With these two terms defined, we can now proceed to a discussion of the increasing importance of enterprise assessment and its role in driving enterprise transformation efforts.

2.2. Introducing Enterprise Assessment

By performing assessment, businesses can understand their current state and use the results to help chart out a transformational path. During the transformation process, assessment provides feedback and a measurement of progress in achieving the desired transformation. Assessment has often been done at several levels of business ranging from specific programs, projects or teams to international enterprises.

Increased globalized competition and rapid innovation have generated a new focus on enterprise-wide transformation (Rouse, 2005), accompanied by a broader, multidimensional measurement of enterprise performance. Rather than simply using bottom-line costs and financial measures (Bourne et al., 2003), enterprises are now approaching transformation with an eye on culture, quality, stakeholder value, leadership commitment, process flow and organizational learning (among others) (Maskell, 1991). As mounting evidence backs this growing array of factors, enterprise assessment becomes both more important and more complex. These intangibles that benefit enterprise performance are harder to measure, because they are not readily apparent in existing data collection strategies (e.g., culture is often not reflected in traditional metrics).

These new enterprise-wide assessments are not designed as replacements for departmental or process evaluation, assessment and transformation, but rather a complement. By examining the broad structural elements and performance indicators of an enterprise, new insights are gained that aid in aligning enterprise actions with sustainable long-term goals beyond the ongoing, short-term financial goals (Kueng, 2000).

The subsequent sections of this chapter introduce four enterprise assessment tools and describe some of the important criteria that can be used to differentiate such tools. We begin by introducing both the challenges and role of enterprise assessment, and then examine four prevalent models used today. With these four models, we examine the different assessment modes, stakeholders, input principles and outputs that tie into the transformation process.

Why Assess

Assessments have long been used for business and operations management. Assessing can act as a leading indicator for shifting performance (for example, "are we retaining our strong performance in these areas?") and for identifying program strengths and/or weaknesses (MIT, 2000). In addition to these internal roles for assessment, it is equally important outside the organization. Assessing can be used to create cross industry comparisons or assist in benchmarking against competitors and standards (such assessments are often used to recognize industry leaders). These results can serve to motivate the organization, boost morale, help complete sales (by sharing assessment results with customers) and even receive recognition or awards. When an enterprise assessment is shared with suppliers, it can be used to align and motivate all companies and players throughout a supply chain to drive a smooth production process and boost output.

In addition, new tools offer enterprise-wide assessments that provide a holistic vantage for identifying the complex interactions of processes and departments across a broad enterprise, bridging the gap between traditional process assessment and enterprise assessment. In order to design, execute and measure an enterprise transformation strategy, having assessments that evaluate multiple dimensions of performance is crucial, both in terms of understanding the current state and charting out the transformation plan. Increasingly, enterprises are trying to

leverage the multifaceted nature of performance in order to gain a competitive edge and maximize value delivery (Burton, Lauridsen, & Obel, 2004; Kueng, 2000). Early stage assessment helps to identify performance gaps and prioritize points of focus, and plays a role in helping to generate a future-state vision for the organization as well as investment precedence. By showing problems with process flow or bottlenecks in organizational performance, assessments provide a key tool in identifying opportunities for improvement. As the transformation plan is implemented, ongoing assessment can then offer feedback and a measurement of progress and return on investment. With this feedback, the transformation plan can be reviewed and revised over time. During enterprise transformation, having a strong and useful assessment tool is crucial in identifying where an organization is, where it wants to go, its rate of progress, and how best to approach that future state (Nightingale & Mize, 2002). And with so much evidence to suggest that simply measuring something leads to improvement (Hauser & Katz, 1998), enterprise assessment ties in to the incentives and motivation for the overarching transformation plan.

Challenges to Assessment

The benefits of enterprise assessment are many, but there are challenges to the assessment process as well. When looking at program or process assessment, it is easier to find assessments that work across companies and industries. But at the scale of enterprise assessment, there is greater variability, both in structure and in values (what should be measured). As a result of this variability, it is significantly more difficult to create a one-size-fits-all tool for assessment. The needs and values of a manufacturing enterprise are quite different than those of a service-oriented enterprise. As a result, enterprise assessments have to make a tradeoff between industry or sector granularity and broad applicability. Even then, enterprises will bypass assessment,

feeling that their unique situation cannot be reflected by a general assessment, but this is often a misconception. Enterprises may have unique challenges, but this does not stop an assessment from being useful; rather than tailoring an assessment tool to the distinctive qualities of the enterprise, the enterprise can use a well-vetted, existing tool and focus on translating the generic outputs into an internally-derived transformation plan and a set of goals, which will reflect the unique nature of the enterprise.

Other challenges result from conflicts between leadership and assessment tools. Enterprise assessment may highlight different foci than the leadership intends, or may be tailored or executed in a way that simply reflects the leadership's desired outcome. Hence an important challenge to good assessment is commitment at multiple levels of the enterprise, both to the assessment tool and its role in the broader transformation process. For the value of an assessment tool to be maximized, an enterprise must commit to using a tool over a period of time (switching tools in the middle of a transformation plan undermines the role that assessment plays in the whole transformation plan), so it is often beneficial to begin assessment on a pilot basis with selected projects before making a larger commitment. Only as familiarity and understanding of an assessment tool grow (as well as historical data), will the enterprise come to fully trust and therefore benefit from assessing. Just as with transformation, assessments must be used to complement leadership (rather than undermine or blindly support leadership values), and require commitment to accurate and on-going usage from those implementing the tool.

A final challenge of assessing an enterprise is the cost or resource allocation required during the assessment process. Either the enterprise must invest in an outside assessment service or key internal personnel must be recruited to participate in the assessment process. This challenge can undermine the near-term support for ongoing or regular assessments. This

challenge is a primary motivator for the creation of abridged assessments, as streamlining the assessment process decreases the resource burden and increases the opportunities and willingness to reassess enterprise performance.

Despite these challenges, assessment has a clear and important role in enterprise transformation and performance improvement, making an investment in assessment worthwhile. One of the ways successful enterprises justify assessments is by making them integral to the transformation cycle (for examples, as the Check step in every Plan-Do-Check-Act cycle). In that light, assessments are a part and parcel of improvement.

2.3. Assessment Tools

With an increasing number of enterprise assessment tools (both public and proprietary), we have opted to bound analysis to publicly available tools that produce numeric performance scores. Such assessments are useful for comparing across assessments and measuring transformation success, and are eligible for the abridged assessment methodology. Of this subset, we will address four different specific assessment tools in this paper: the Malcolm Baldrige National Quality Award, the Good to Great Diagnostic, the Lean Enterprise Self-Assessment Tool and the Shingo Prize for Operational Excellence. Each is introduced below along with relevant background. Two of the assessment tools are also business recognition awards, which have widely published their assessment criteria and process. These are included here because organizations have been able to adapt published criteria to create internal assessment tools (Abdimomunova, 2010). Of course, such award-based assessment tools have added incentives for implementation. In addition to measuring and assessing transformation or other internal processes, the enterprise focuses on improving qualities that can draw in national or international recognition.

2.3.1. Baldrige Prize

The Malcolm Baldrige National Quality Award is a program funded by the US Department of Commerce to recognize organizations that demonstrate performance excellence, as well as world-class product or service quality (ASQC, 2009). The program has been managed by the National Institute of Standards and Technology (NIST) since its inception in 1987. The program has been viewed as successful in motivating and driving quality and performance values, with net private benefits to the economy estimated at over $24 billion (Link & Scott, 2006).

The program has a clearly defined and published set of criteria for the Baldrige award, which includes scoring guidelines and self-analysis tools (such as preliminary and supplementary self-assessments). Although designed for evaluating and awarding the national prize, the criteria reflect a range of important organizational performance indicators that are used to drive a transformation plan. The assessment involves mapping out key processes and answering qualitative questions regarding organizational strategies and practices; these questions are then scored against clearly defined criteria. Each score can then be combined to a total score, up to 1,000 points.

The Baldrige criteria provide a thorough, quality and performance driven assessment that can be performed internally. It's applicable to many sectors, and has complementary criteria sets specific to certain sectors and sizes (including healthcare, education, manufacturing, and small businesses).

2.3.2. Good to Great Diagnostic

The Good to Great Diagnostic tool is based on the book *Good to Great: Why Some Companies Make the Leap… and Others Don't* by Jim Collins (2001). In this book, 1,435 good

companies are examined to identify those that made a substantial and sustainable increase in performance (a sudden inflection point). Of those that made this performance shift, 11 were selected and examined to understand what indicators and common elements may have influenced their exceptional performance. From this analysis, a book of observations and best practices was created as a guide for achieving "greatness" in companies (Collins, 2001).

Along with the book, Jim Collins developed a diagnostic tool for differentiating between "good" and "great". The assessment includes ten key inputs (concepts that need to be implemented) and 3 key outputs (indicators) to measure both the implementation of the concepts identified in the book as well as trends in their implementation.

The tool is broadly applicable (the 11 companies used were from a range of industries and sectors, and focuses on the common, cross-sector best practices). Overall, it examines quality leadership and commitment to core values, with a willingness to approach all others ideas with flexibility and honesty.

The Good to Great Diagnostic is easily implementable and provides a summary of some top level, broadly assessable indicators that are associated with and demonstrated by companies that are able to achieve strong growth patterns over a substantial amount of time.

2.3.3. Lean Enterprise Self-Assessment Tool

LESAT (Lean Enterprise Self-Assessment Tool) is a tool developed by a team of industry, government and academia members brought together and facilitated by the Lean Advancement Initiative at MIT. It is a questionnaire designed to be used as a self-assessment, involving the top leadership of an organization. The tool was originally designed to fit into an existing transformation plan – the TTL (Transition to Lean) Roadmap (MIT, 2000), which has since been replaced with the Enterprise Transformation Roadmap (Nightingale, 2009) – but

works independently of these frameworks. LESAT includes leading indicators associated with organizational "leanness" (a broad business paradigm that has been gaining interest and adoption, and is based on principles identified in the success of the Toyota Production System (Womack, Jones, & Roos, 1990)) and prioritizing/assessing gaps between the current state and a desired state (Nightingale & Mize, 2002). Because of the tool's dual purpose in measuring the current state and envisioning a future state, it has substantial value from a transformation perspective.

The tool was originally developed with input from the aerospace industry, yet it has substantial applicability for manufacturing enterprises in a diverse range of industries. Increasingly, the tool is being used in healthcare and service industries. LESAT is notable in its role as self-assessment, integrating many perspectives and vantage points into an assessment that can drive transformation by identifying performance gaps and measuring improvements in the implementation of lean enterprise principles.

It differs from the Baldrige Award in that it is based on a specific paradigm (lean) and it explicitly evaluates performance relative to lean practices. Because Baldrige was developed through an aggregation of best practices, its criteria reflect some lean concepts. A mapping of practices between the two assessments identified 17 practices in LESAT that are not specifically addressed in the Baldrige Award (MIT, 2001).

LESAT will be discussed again in subsequent chapters, as it is well-suited to the suggested methodology: the high number of inputs and its intended use as a self-assessment create high resource requirements, as many enterprise leaders must devote substantial time to the assessment process. Creating an abridged version of LESAT will help decrease the associated investment and enable more regular reassessment.

2.3.4. Shingo Prize for Operational Excellence

The School of Business at Utah State University awards the Shingo Prize for Operational Excellence annually to companies that achieve world-class manufacturing (Shingo Prize for Operational Excellence, 2009). It was started in 1988 and named after Shigeo Shingo, a Japanese industrial engineer who played an important role in designing the Toyota Production System. As a result, the Shingo Prize reflects a number of similar criteria addressed in LESAT, including lean implementation and value creation.

The Shingo prize addresses a range of dimensions, including organizational culture, continuous process improvement and business results. Scores are assigned across a number of dimensions and these are combined into a weighted overall score, which – if above a certain threshold – corresponds with an achievement level of Bronze, Silver or Gold.

The Shingo prize has its origins and focus in manufacturing, but has recently been revised to offer a broader range of performance indicators. The shift in focus is mirrored in its name, having previously been called the "Shingo Prize for Excellence in Manufacturing". The Shingo Prize has many conceptual similarities to both the Baldrige Award and LESAT. Like the Baldrige Award, it seeks to balance several proven best practices, ranging from personnel development to financial success. Like LESAT, Shingo Prize has its origins in lean concepts (via the Toyota Production System). These complementary concepts are combined in a unique framework describing the hierarchical progression of organizational performance, moving from cultural enablers up to business results.

2.4. Comparison of Enterprise Assessment Tools

The four assessment models described (Baldrige Quality Award, Good to Great Diagnostic, LESAT and the Shingo Prize) are all strong assessment tools that can provide the much needed data to drive enterprise performance and/or enterprise transformation.

Yet, each tool has different strengths and the choice of assessment tool can be influenced by a variety of criteria. The following criteria are discussed in terms of the four models listed, but can easily be applied to any assessment tool. We have selected our analysis criteria to include: (1) the causative or correlative link with improved outcomes, (2) the mode of assessment, (3) the assessment stakeholders, (4) assessment inputs or criteria, and (5) assessment outputs or information gleaned. Collectively, these criteria provide an overarching understanding of the assessment tool and it use. By addressing and understanding the criteria listed in this section, selecting the most appropriate assessment tool will be easier. The aim is not to identify one assessment as superior, but instead to simply show differences in focus, scope and implementation, so that readers understand the variety and structure of enterprise assessment tools that exist and are used.

2.4.1. Causality

The ideal criterion for picking an assessment tool is the evidence establishing a causal link between actions taken to improve assessment scores and an improvement in enterprise performance (i.e., as a company makes changes to improve its assessment score, these changes result in improved performance). Unfortunately, this sort of causal link has not been established and isolating such a causal link is almost impossible. The reason for this gap in understanding has to do with the challenges associated both with isolating the factors that influence performance and with conducting enterprise experiments. Enterprise performance is

multidimensional, influenced by a range of factors (e.g., climate, culture, strategy, investment, quality, innovation, etc), where many of these factors have complex interactions (i.e., two independent qualities associated with success may combine to become a deterrent) and time delays (Burton et al., 2004). As a result of the complex, systemic nature of performance, the direct causal influence of assessment principles on increased performance is difficult to tease out of available data. To further complicate the analysis, implementation and utilization of assessment tools varies substantially between organizations, so assessment scores may not be consistent in their representation of corrective action.

Most assessments are built from correlative data or demonstrated enterprise models. As a result, the assessment looks at principles or strategies that have a strong correlation with successful enterprises. For example, the "Good to Great Diagnostic" looks for common cultural elements and business practices among exceptional companies. But the data to show that these principles or strategies *cause* the success do not exist (van De Ven & Ferry, 1980).

Although understanding this causal link would make evaluating different assessment methods much more concrete, we instead must accept the lack of solid data and look at different criteria. Assessment tools should be evaluated based on their alignment with the enterprise mission and values. Even though the assessed variables cannot be shown directly to increase performance, they will drive the core enterprise values and help ensure that the enterprise reflects those values. And once selected, an enterprise should commit to the assessment tool, as well as the broader transformation process, to maximize the value derived from ongoing assessment and transformation. Too often organizations switch assessment tools thinking that an alternative tool improves performance more. But as described, such causal data does not exist and switching might undermine value alignment and commitment achieved by the first assessment.

Selecting an assessment tool that reflects core mission values and receives buy-in from all parties is crucial to maximizing the benefits from assessment. As experience and understanding of the assessment tool grows, results and performance for the core values will also grow (as mentioned earlier, evidence suggests that measuring certain criteria will naturally result in increased performance) (Hauser & Katz, 1998). Therefore, despite the lack of causal evidence associating an assessment method with performance, commitment and value alignment improve the value gain from assessment.

In order to compensate for the inability to show crucial causation data, research has been done to compare model reliability and validity based on correlations. Internal correlations (looking at cross-correlations in assessment criteria) can be used as an indicator of strong model reliability, since the criteria are all measuring a certain common factor (in this case, a holistic view of enterprise performance), assuming it exists (Furr & Bacharach, 2008). Other studies have examined the progression of assessment results over time as evidence of model validity (Hallam, 2003). These correlation studies help fill in some of the questions left unanswered by the lack of causal data by investigating the internal validity and reliability of assessment tools.

2.4.2. Assessment Modes

Enterprise assessment can be performed a number of different ways. Broadly grouped, there are three primary modes. Different assessments lend themselves better to different implementations, but can easily be adapted to another usage. In addition, each usage carries distinct advantages and disadvantages which may influence both the assessment used and how it is implemented.

Managed or External Assessment. Both with proprietary assessment tools, and with the award/prize-based assessments, the intention is to have a third-party perform some of the data collection and compilation of assessment results. In these cases, assessment experts who are familiar with the assessment model can be hired to perform the actual assessment. This can result in objectivity, consistency and reliability (van De Ven & Ferry, 1980) because of the advantage in terms of tool knowledge. This knowledge helps make sure the assessment is performed in a manner consistent with its intent, and can even be used to tailor the assessment tool to the needs of the enterprise. External assessment provides opportunities for benchmarking across multiple companies or industries, which is sometimes worthwhile for organizations interested in exploring best practices.

Of course, the high familiarity with the tool comes with an important trade-off: the assessor may have limited knowledge of the enterprise being assessed. This internal enterprise knowledge can help guide the assessment process and ensure that all relevant information is reflected in the assessment. In addition, managed assessment can come with a high price tag.

Due to the steep learning curve associated with some tools, managed assessments are beneficial for performing one-time assessments. Managed assessors can also play a role in introducing a new assessment tool, by customizing it, training relevant staff and assisting with the early implementation of the assessment (Abdimomunova, 2010).

Hybrid Managed/Self-Assessment. Organizations increasingly are creating or fostering independent internal entities to drive, measure and encourage transformation. Assessments often fall under the purview of such an entity. Assessors within the assessment group can gain a high level of familiarity with both the tool and the enterprise structure (Abdimomunova, 2010). As a result, the assessment may be well performed.

Internal divisions may be affected by leadership pressure or bias, but still have the advantage of being a distinct department that allows for a more objective view of processes or performance in other departments. Despite the still rather high costs (maintaining personnel and departmental resources), this approach allows for strong tool and organizational familiarity, along with better data management in terms of tracking or comparing assessment trends over time. In addition, such hybrid models allow for assessment to be fully integrated into the natural business cycle, decreasing the negative impacts or the obstructive nature of outside assessment.

Self-Assessment. The third approach to assessment involves self-assessment, where many internal stakeholders or parties take part in the assessment of their own performance. This introduces a broader range of perspectives, and greater familiarity with day-to-day processes and its own set of potential challenges or problems (Nightingale & Mize, 2002). Case studies suggest self-assessments can be more time intensive, due to the added steps of coordinating and facilitating the assessment process, collecting feedback and reaching consensus among multiple participants (Abdimomunova, 2010). This high time commitment makes self-assessments especially suited to being abridged using the methodology introduced in subsequent chapters. Despite having a higher time commitment, the process is often less costly than managed assessment or having a dedicated internal department, since there are less external personnel requirements.

The downside to self-assessment is the high knowledge prerequisite for completing the assessment. Participants must be familiar with the model and the assessment tool. As a result, self-assessment can sometimes be self-selecting by only getting feedback from those most familiar with the model (Abdimomunova, 2010); potential strategies exist for overcoming this through facilitation and dedicated, collocated assessment time blocks. Also, self-assessment can

result in less consistent results or interpretation of principles. Finally, responses can reflect different sorts of biases (appeasing superiors, over scoring oneself, etc) (van De Ven & Ferry, 1980).

As the low familiarity obstacle is overcome by education borne through a long-term commitment to the self-assessment strategy, a key advantage of self-assessment appears: greater conversation about assessment principles. As many stakeholders are involved in the assessment process and are brought together in order to discuss diverging scores, the tool can inspire and encourage conversation and cross talk among different personnel, departments and stakeholders.

The three different modes of assessment and their associated tradeoffs are summarized in Table 1.

Table 1 – Modes of assessment and the associated tradeoffs

	Managed/External	**Hybrid**	**Self-Assessment**
Tool/Model Knowledge	Extensive	Extensive	Limited
Org. Knowledge	Low	Medium	High
Costs	High	Medium	Low
Time	Low	Low	High
Bias	Low	Medium	High
Best Uses	One time assessments or introducing a new tool or exploring best practices	Ideal for long-term, high-commitment transformation plans that involve regular assessment and data analysis	Good for reflecting detail and a variety of vantage points and encouraging conversation or involvement in the transformation process

2.4.3. Assessment Stakeholders

Different assessment tools are designed to have different levels of granularity in terms of who is included in the assessment process. In many cases, assessments seek to get the input of individuals who have the best ability to observe and measure the assessment's critical values and indicators. Often, the assessments can easily be tweaked or modified to change the involvement

of the assessment to include more or fewer participants, depending on the goals of the assessment.

In addition, the scope of the principles and results of assessment tools vary. Some have a broad scope that looks at a range of enterprise indicators and performance metrics both internally and across stakeholders, while others focus in on a core set of internal indicators.

Depending on the goals of the assessment and transformation plans, both the personnel involvement and the assessment scope are important factors in selecting the right tool.

Malcolm Baldrige National Quality Award. The Baldrige assessment process is flexible in its personnel involvement. Since it was originally designed as a prize application, there are few details regarding who should be involved in the assessment process (since all assessments are verified by an independent examiner). When used simply as an internal assessment tool, it can be divided among many people to map out processes specific to their contribution.

The overall scope of the assessment is a high level assessment, with a large leadership focus and substantial attention to the customer/end-user (in terms of quality product or service delivery).

Good to Great Diagnostic. The "Good to Great Diagnostic" is designed to be used by top leadership, who have a comprehensive perspective of company dynamics. The diagnostic tool focuses substantially on personnel, leadership and culture across an organization, but also looks outside the organization to community and shareholder values.

Lean Enterprise Self-Assessment Tool. LESAT is a self-assessment tool, and therefore looks for many perspectives and participants. It specifically targets enterprise leadership, who are in touch with the overarching enterprise goals and performance. Since the tool is designed to

assess the enterprise as a whole, rather than individual departments, this global perspective is important. The one limiting factor on participation is the fact that the tool requires a substantial level of knowledge regarding lean principles and concepts.

LESAT has a broad scope and looks at the larger context of the enterprise as well. It can easily be implemented or applied across organization boundaries in order to ensure lean principles flow throughout the extended enterprise.

Shingo Prize for Operational Excellence. As with the Baldrige Quality Award, the Shingo Prize was originally designed for evaluating award applicants. As a result, using Shingo Prize as an assessment tool requires some modification, but has a large amount of flexibility in terms of the personnel involvement.

2.4.4. Criteria or Information Assessed

One of the most important factors in deciding on an assessment model is the information or criteria measured by the assessment. For an assessment tool to succeed and achieve the necessary organizational buy-in, it must measure those principles and criteria that most closely align with either the enterprise values and/or the goals of the transformation plan.

In addition, the actual scoring mechanism is important to assessment, in terms of achieving the desired level of granularity and information specificity necessary to monitor and guide change.

Malcolm Baldrige National Quality Award. The Baldrige Award considers seven different broad performance measures: leadership; strategic planning; customer focus; measurement, analysis and improvement of organizational performance; workforce focus; process management; and results (ASQC, 2009). Each of these categories has many questions/prompts meant to elicit details regarding the processes, implementations and success of

that specific topic. Based on defined scoring levels and criteria, the qualitative responses are converted into a percentage score.

Good to Great Diagnostic. The diagnostic tool measures the implementation of five key concept groups, each with two practices described in the *Good to Great* book (Collins, 2001). The criteria focus around leadership ("disciplined people"), organizational culture ("disciplined thought" and "disciplined action"), and sustainability ("building greatness to last"). In addition, the diagnostic looks at three outputs that are intended to provide an objective measure of the concept implementation: performance superior to mission, distinctive impact on community, and lasting endurance beyond significant changes.

Overall, there are about one hundred scores assigned, on a grade scale (A-F), as well as trend scores (improving or declining performance). These can be aggregated to measure the success of the five key concepts outlined in Collins's book.

Lean Enterprise Self-Assessment Tool. The Lean Self-Assessment Tool has 54 practices broken into three key categories: lean transformation/leadership, life-cycle processes and enabling infrastructure (Nightingale & Mize, 2002). Primarily, the practices and assessment criteria are closely tied to the predominant literature on lean enterprise principles (Murman et al., 2002) and reflect knowledge gained from over 27 theses generated by the Lean Advancement Initiative.

Each principle is scored on a five level scale, meant to provide the level of "leanness" for each specific criterion. Principles also receive a desired score based on a predetermined transformation timeline, which helps in crafting a future vision for the organization and in identifying high-priority areas for transformation.

Shingo Prize for Operational Excellence. The Shingo Prize evaluates 17 key principles, evaluating different levels of commitment to the principles. The principles are primarily centered around concepts from the Toyota Production System and lean business practices (Shingo Prize for Operational Excellence, 2009). Principles include cultural enablers (leadership, ethics and personnel development), continuous process improvement (lean ideas, value stream and support processes), consistent lean enterprise culture (enterprise thinking and policy deployment) and business results (creating value).

The scope of the assessment is broad, encompassing individual processes, safety, organizational culture and lean thinking. The assessment is defined progressively to measure different levels of innovation and principle adoption, looking first at a tool level, then a system level and finally a principle level. This tiered model makes Shingo Prize a comprehensive assessment of the interconnected dynamics of organizational performance. In addition, some stakeholder values are addressed, in terms of financial results and quality levels.

Each principle has scoring criteria, and is given a weighted score. These scores can be combined into an overall score that is representative of an organization's operational excellence.

2.4.5. Information Gleaned

The final important element of the assessment process is the output; an assessment tool must provide a usable, understandable output that can guide the transformation process. Some tools also provide overall scores, which can be beneficial in providing cross-industry benchmarks or providing a marketable indicator of success. Important to picking an assessment tool is its ability to output useful, functional results that can guide and measure the transformation process.

With the exception of the Good to Great Diagnostic, the assessment tools reviewed in this paper do not contain explicit interpretation guidelines. Although the models all describe the desired states, most do not have instructions or transformation guidelines that map out the course to move up a level for each performance indicator.

Malcolm Baldrige National Quality Award. The Baldrige Prize provides general criteria for each level of performance, which can guide the transformation process. In addition, they have published worksheets and strategies for identifying opportunities for improvement as well as strengths (further growth), which help in formulating a future state vision while completing the assessment. In addition, the assessment questions include details about actionable items and responsible parties in order to help develop implementable, achievable changes.

Overall, the model encourages working out from the core values to develop systematic processes that yield beneficial results. This concept provides a guide to connecting internal processes with stakeholder value to achieve high performance.

In addition, the assessment process generates an overall score that can be used to measure organizational improvement and used as an inter-organization benchmarking tool.

Good to Great Diagnostic. The Good to Great Diagnostic links heavily with the principles and concepts described in Collins's book on organizational greatness. The scores that come from the diagnostic tool can be used to focus and foster implementation of different concepts from the book (essentially, identifying best practices that can benefit the enterprise being assessed). In this case, the assessment tool is very closely tied with a specific set of principles and an associated transformation plan.

Lean Enterprise Self-Assessment Tool. LESAT yields two sets of scores, one for the current enterprise state and one for the future/desired state. Rather than combining the individual

scores into an overall, unified score, each is used independently to assess current performance vs. the desired state. Often, the results are analyzed by identifying performance gaps (the difference between the current state and the desired state) in order to identify key priority areas for transformation (Nightingale & Mize, 2002). The assessment can later be repeated to measure progress towards achieving the desired state, as well as for measuring organizational focus changes.

In addition, as a self-assessment, LESAT provides interesting results in terms of variability or non-convergent responses. By looking at these variations, organizations can identify perspective difference and misalignment across departments or management levels. The next chapter focuses on LESAT specifically, and provides more information on the assessment outputs and their interpretations.

The LESAT was originally designed as a component of the Transition-to-Lean (TTL) roadmap, and its outputs map to key steps in the transformation roadmap. Yet it is a sufficiently flexible assessment that it can easily be adapted to other transformation frameworks.

Shingo Prize for Operational Excellence. The Shingo Prize generates individual scores for each principle, as well as a single unified score. The unified score can be used as a motivational force (achieving scores associated with "gold" operational excellence), while the principle scores can be used to define the transformation focus. Each principle is accompanied by concepts and criteria to help envision the desired state.

Although the Shingo Prize does not integrate with an implementation strategy, it provides a framework for prioritizing improvement and envisioning a future state. In addition, a unique progression of practices helps enumerate the phases of transformation and adoption (starting at the level of tool-driven change and moving up to principle-driven change).

2.5. Summary

The importance of enterprise assessment cannot be overstated, though rarely practiced. As enterprise transformation becomes increasingly important to remain competitive, the ability to identify, measure and integrate a multitude of performance indicators is valuable. It provides both a measurement of the enterprise's current position within the industry, as well as prioritized focus for how to improve performance.

Because it is quite difficult to demonstrate causality between improving assessment indicators and improving overall enterprise performance, assessments must be chosen based on a range of other qualities. Most important to the success of enterprise assessment is commitment and buy-in, both at the level of leadership and of implementation. Any tool that has a high level of buy-in and aligns with the enterprise values and goals will be beneficial to the transformation process. A well-aligned tool, if integrated into strategic processes, is able to provide beneficial knowledge about organization performance and priorities, as well as historical data and insights into the multidimensional factors that contribute to success.

This chapter has provided a cursory introduction to four leading enterprise assessment tools, and outlined differences in their scope, criteria and outputs (Table 2 below summarizes this information). Of course, the tools outlined in this chapter are a sampling of those available. More tools are constantly appearing, and existing tools can easily be customized to better align with an enterprise's focus and goals. By introducing the select number of assessment tools, this chapter has sought to ground subsequent discussions in the larger realm of enterprise assessment. With increasing need for enterprise assessment and with the increasing pressure of resource constraints, it is imperative that flexible and agile tools exist to enable the enterprise realize assessment benefits on an ongoing basis. The abridged assessment methodology defined in

Chapter 4 seeks to reduce the resource burden of assessment by developing complementary versions of enterprise assessments with fewer input criteria.

Table 2 – Summary of four enterprise assessment models

	Baldrige Prize	**Good to Great**	**LESAT**	**Shingo Prize**
Assessment Mode	Award, can be adapted for internal assessment	Internal diagnostic to distinguish between good and great companies	Self assessment in support of transformation planning	Award, can be adapted for internal assessment
Assessment Stakeholders	Flexible	Top leadership	Enterprise Leadership	Flexible
Criteria or Information Addressed	Quality and customer commitment	Best principles identified in *Good to Great* book	Lean enterprise practices	Toyota Production System and lean manufacturing
Information Gleaned	Areas for improvement and key principles	Trends in implementation of concepts	Gaps and prioritized improvement areas	Successive adoption pyramid guides transformation
Sectors	Manufacturing, service, small-business, health, education, non-profit	Broad	Designed for manufacturing (aerospace), recently applied to healthcare and services.	Designed for manufacturing, recently expanded to Operational Excellence.
Sample size (approximate)	Thousands	Based on 11 Usage unknown	Dozens	Hundreds

Chapter 3. The Assessment Process and Insights: LESAT

Having now established the role and importance of enterprise assessment, this chapter will focus on specific tool's process for implementation and for interpretation. Understanding the detailed process of assessment and how insights are derived from the assessment results is crucial to the construction of an abridged assessment, as it is important that the abridged assessment still retains the key values of the assessment process (such as defining the scope, promoting discussion and embedding transformation knowledge in participants) as well as the value of the interpretation process (such as tracking enterprise progress, tracking personnel cohesion and driving enterprise).

This chapter focuses on LESAT – the Lean Enterprise Self-Assessment Tool – which was introduced in the previous chapter. The tool was developed by a team of industry, government and academia members brought together and facilitated by the Lean Advancement Initiative at MIT. It is a questionnaire designed to be used as a self-assessment, involving a range of employees, managers and/or leaders within an organization. The tool assesses 54 practices in three sections: leadership and lean transformation, life-cycle processes and enabling infrastructure (the full list of practices appears in Appendix A: LESAT Structure). Each respondent scores each practice, providing both a current state and a desired state score. The assessment process involves facilitation to introduce the tool and practices, coordination of key enterprise leaders in order to complete the assessment and discussion of results and implications. The tools outputs shed light on a range of practices and their relative impact on enterprise performance, as well as insights into trends and group cohesion.

As a self-assessment, LESAT is well suited to the abridged assessment methodology. Due to the high number of practices and the high number of respondents involved, the time and

personnel commitment required is substantial. This has been confirmed during interviews with past and current users who have commented that the high time commitment has acted as a disincentive in on-going usage, especially when it comes to performing reassessments (Abdimomunova, 2010).

3.1. Assessment Process

The prescriptive assessment process for LESAT originally appeared in the facilitator's guide (MIT, 2001), but has since been revised by research emerging from the Lean Advancement Initiative (Abdimomunova, 2010). The multistep process provides guidelines on facilitating the assessment, scoping the enterprise, introducing the tool and completing the scoring. Having a clearly defined process helps overcome threats to validity and ensures that the enterprise is able achieve a systematic, consistent and reliable assessment that helps align enterprise leaders and generates actionable insights.

In this section, we briefly review the recommended assessment process as established by Abdimomunova (2010) during her research with the Lean Advancement Initiative. The process is based on five case studies of LESAT implementations at four different aerospace organizations. From the case studies, feedback and literature, four guiding principles were derived.

3.1.1. Guiding Principles for Assessment Process

First, there should be a shared understanding of the objectives and benefits from the assessment process by all stakeholders (such as leadership, facilitators, participants and respondents). By fostering common objectives and goals, there is both greater motivation and greater buy-in for the assessment process, laying the groundwork for a successful

implementation and for continued commitment. This is achieved through collaborative planning, continuous communications and iterative refinement of goals and objectives.

The second guiding principle is that there should be consistent knowledge of the underlying principles being assessed. Having sufficient knowledge of the principles is important to the accuracy and usability of the assessment results, but it also is important for establishing investment in the assessment process. Respondents and participants will see more value in the assessment process if it is a topic that they understand. Using training, either specifically for the assessment participants or generally for personnel, helps create the knowledge to understand both the principles being assessed and to see their long-term viability in terms of potential payoffs for the enterprise.

Third, the assessment process must be accompanied by frequent communication. Although frequent and clear communication is implicit in the last two guiding principles, it is an important overall tenant of the assessment process. Stakeholders, facilitators and participants must maintain frequent and consistent communication in order to reinforce the objectives and benefits of assessment, to encourage buy-in and commitment to the process, to create both a sense of transparency and appreciation for the process and, finally, to ensure that the assessment results translate into enterprise behavior and actionable projects. Similarly, communication must extend outside the boundary of the assessment team to the whole enterprise. The assessment process should not be performed by a small leadership team behind the closed doors, but should be a transparent process shared with the full enterprise as outcomes will come to affect the whole enterprise. Communication to the whole enterprise should include the motivation, the process, the values inherent in the assessment, any opportunities for staff input (either via a relevant leader who is involved in the process or another avenue) and finally the outcomes and their

implications. Communication is a crucial guiding principle to the assessment process as it helps create a common vision for the process and to ensure that it aligns with the needs and expectations of the enterprise.

Building on communication, the fourth guiding principle for the assessment process is open discussion. For assessment to be valuable, participants must feel that they have an opportunity to speak openly and provide honest feedback for all practices assessed. This helps the process to reach a true consensus, and allows the assessment process to reach creative and nonobvious insights regarding the transformation plan and potential obstacles. Communication should occur both during the assessment process, but also during the results session as participants may gain value from sharing their rational for certain scores, ideas for improvement or uncertainty about key terminology. All this feedback can help drive new learning opportunities, discussions of core concepts and reevaluations of the functions and roles of individuals and departments within the enterprise.

3.1.2. Recommended Assessment Process

These four guiding principles serve as a backdrop for the five-phase assessment process depicted in Figure 1 (Abdimomunova, 2010). The process begins with establishing assessment prerequisites, such as the appropriate resources and the needed enterprise commitment and consensus regarding the objectives for the process. The second phase is a planning phase, where the prescriptive process is adapted to the local enterprise and its unique needs. During the third phase, the assessment is actually administered by collecting and analyzing results (this is the "measurement" part of the assessment process). Next, the fourth phase is evaluating the results and the process in order to improve both the enterprise and to refine the assessment process for the future (this is the "evaluation" part of the assessment process). Finally, the fifth phase of the

assessment process closes the assessment loop by integrating assessment results and insights into the enterprise strategic plan. During the final phase, the outcomes and results of the assessment are translated into opportunities for improving enterprise performance. These five phases are sequentially linked, with the output from each phase acting as an input to subsequent phases. In addition, the process involves feedback loops where the process is repeated and adjusted based on lessons learned during previous iterations.

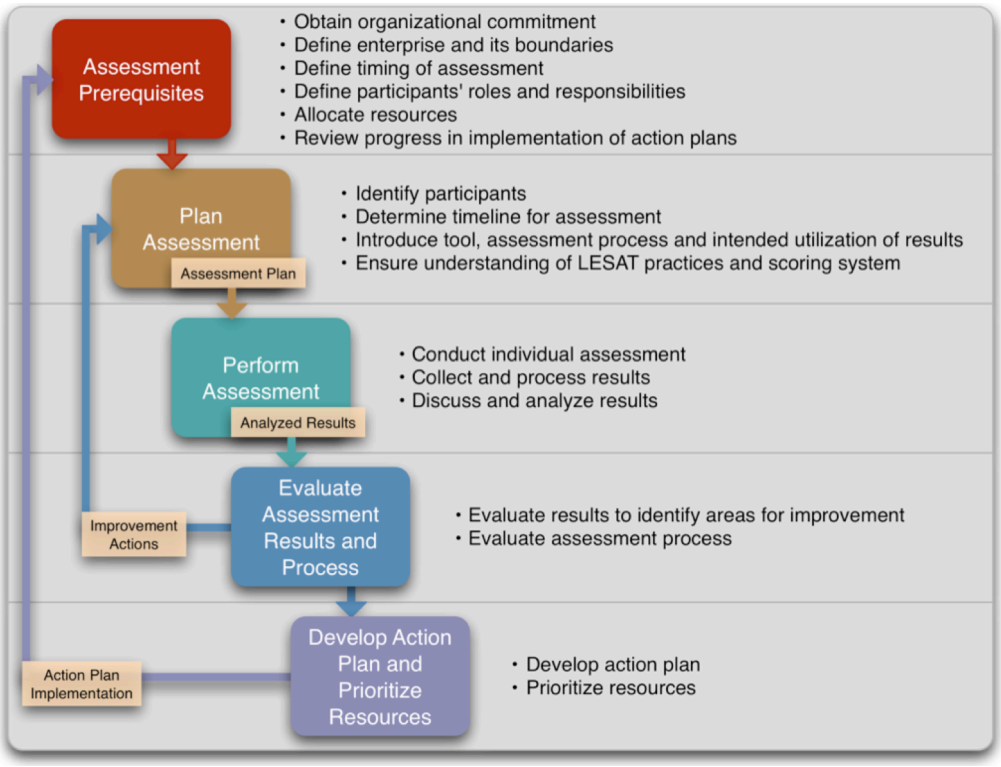


Figure 1 – Recommended LESAT process (Abdimomunova, 2010)

3.1.3. Assessment Stakeholders

Before actually describing the individual phases, it is worth establishing explicit definitions of the groups of stakeholders involved in throughout the process. There are four key groups of stakeholders who are integral to the assessment process. Although there is overlap

between the different groups of stakeholders, each has unique responsibilities and involvement in the assessment process.

First, the *enterprise leadership group* refers to individuals whose performance and responsibility are at the enterprise level. The leadership has an important role in selecting the assessment, demonstrating commitment both to the process and the results, and allocating the necessary resources for the assessment.

Next, the *respondents* are those individuals who are actually scoring the enterprise on all practices. Based on the design of LESAT (Nightingale & Mize, 2002) and the enterprise scope of the practices, respondents should all have a full enterprise perspective and be able to assess the whole enterprise on each practice. Generally, respondents will include members of the enterprise leadership.

Third, the *assessment facilitator* has an important role overseeing the whole assessment process and providing a coherent environment that integrates each phase of the process. Although the facilitator is not a fulltime task, she/he should have ample time to devote during the key assessment period and it should be a dedicated responsibility (ensuring consistency in the facilitation during each cycle of the assessment process). The facilitator responsibilities including coordinating the different groups of stakeholders, fostering commitment, instigating communication, conducting the assessment, performs analysis, etc.

Finally, the *assessment users* represent those who will benefit from the outputs of the assessment process and who will be acting on the assessment results. Broadly speaking, this can includes managers, employees, leadership and more (stakeholders in the three other groups may also constitute assessment users). Although this group may not directly interact with each phase of the assessment process, they must feature prominently in the planning and communication

components of the process. That is, the plan must reflect the needs of the assessment users and the assessment process must be well communicated and transparent to the assessment users in order to win their support and commitment to the results.

3.1.4. Phase One: Assessment Prerequisites

The first phase of the assessment process is to foster the appropriate environment for the assessment, ensuring that the enterprise conducts a beneficial and actionable assessment with relevant results. Because of the complexity of proper assessment – requiring long-term commitment, strategic alignment, iterative assessments and careful analysis – it is important that a sufficiently supportive environment is established before embarking on the assessment process. This is achieved through a range of initiatives:

Enterprise Commitment. The first, and one of the most important, is to obtain enterprise commitment. The assessment process must be prioritized and committed to by enterprise leadership. Organization commitment fosters a sense of legitimacy among respondents, generating greater engagement and more valuable feedback. To achieve enterprise commitment, leadership must understand and value the benefits from assessment (as described subsequently in section 3.3).

Define the Enterprise and its Boundary. Setting a clear scope and boundary for the assessment process will enable facilitators to better plan for the assessment and will enable respondents to better measure enterprise performance. The established boundaries have implications outside of the assessment process, effecting the transformation planning process too. For this reason, the scoping of the assessment should be done simultaneously with the drawing of boundaries that focuses the transformation process. To properly define the scope of the enterprise, the definition presented earlier must be considered (see the definitions in section

2.1). Broadly speaking, the extended enterprise is a full value stream from creation of value to delivery of value, all related stakeholders and the supporting processes that enable the value creation. For the purposes of the assessment process and transformation plan, a more limited definition of the enterprise is used based on a reasonable sphere of control (only assessing and transforming those components which fall within the current organization).

Define Timing of the Assessment. A broad assessment timeline must be established early in the assessment process in order to properly link assessment results with the strategic planning and transformation planning needs. The LESAT results are designed to provide both insights about the current performance of the enterprise, as well as aid in prioritizing transformation efforts. In order for these benefits to be realized, LESAT results must be available sufficiently early in the planning process. Enterprise leaders should coordinate with the assessment facilitator to set an overarching timeline for the initial assessment, as well as reassessment dates, so that pertinent and current data is available during crucial strategic planning sessions.

Define Participants' Roles and Responsibilities. Early identify of the facilitator, respondents and users is also important for the planning process, as it allows key groups to be involved in planning the assessment and ensure they are involved and invested in the process by allowing them to contribute to the planning. Enterprise leaders should select the LESAT facilitator, and then the leaders and facilitators should work together identify assessments users.

Early identification of the facilitator provides an ample opportunity for pre-assessment training, which is especially important for the assessment facilitator. In LESAT's case, this implies that participants should be broadly familiar with the definition of a lean enterprise, related terminology, lean enterprise principles and the assessment tool itself.

Allocate Resources. Since LESAT is designed for internal facilitation and is a self-assessment, the primary resources needs are people and time. All participants must be able to commit sufficient time to complete the assessment and attend introduction/results meetings. For this reason, it is fundamental that the enterprise leadership show support for the process and recognize individual involvement in the process as value-added.

In addition, resources should be tentatively allocated for implementing projects based on the results of LESAT. Early allocation of resources for transformation projects is an added way that enterprise leadership can demonstrate their commitment to the assessment process.

Review Progress in Implementation of Action Plans. This final step involves reintegrating previous lessons learned from past assessment cycles into the planning process. This is two-fold: both a review of the lessons learned about the assessment process as well as a review of the actionable projects that results from the past assessment. This review step will help frame the second and all subsequent assessment processes in the context of the enterprise. For example, if the past actionable projects were quickly implemented and realized, such a review may boost confidence that more ambitious goals can be set during this assessment cycle.

Having completed the six initiatives described above, the enterprises should be primed to plan the assessment process.

3.1.5. Phase Two: Plan Assessment

Now that the appropriate assessment environment has been established, the enterprise can begin planning the execution of the actual assessment.

Identify Respondents. The first step is to identify respondents who will score all the LESAT practices. The facilitator should work with enterprise leaders to identify potential respondents (both internal and external to the leadership group). As a self-assessment, the results

benefit from a diversity of views and inputs (it both reduces statistical error and allows group cohesion to be better understood), but all respondents should be individuals with a perspective of the full enterprise (as they are being asked to score overall enterprise behavior) and have personal responsibility for at least a portion of the enterprise. There is no "right" number of respondents, but at least five helps reduce statistical error and groups over thirty become unwieldy and provide little benefit in terms of precision (Abdimomunova, 2010; Hubbard, 2007). If the enterprise intends to compare results across subgroups (i.e., different levels of management or different departmental perspectives), then each subgroup should have at least five respondents. This subgroup comparison enables more in-depth analysis and can drive more localized responses.

During subsequent reassessment, using the same respondents is always preferable as it provides better cross-time comparison of results. This will also allow the assessment process to capitalize on existing knowledge and familiarity with the assessment tool.

Determine Timeline. Once respondents have been identified, the facilitator can begin setting a specific assessment timeline based on the overarching strategic timeline previously established. The timeline should be set with input from respondents and assessment users, as the process will effect them and should not be viewed as an external burden. Setting a timeline early is important as LESAT requires attention to detail, ample time for discussion and the attention of busy, high-level leaders.

Generally, the timeline spans three meetings on consecutive days. Ideally, each meeting should span between half and a full day each. The first meeting covers the training and introduction of the assessment tool, followed by the actual scoring of practices. The facilitator collects the responses and prepares initial analysis for the second meeting, when results are

presented and discussed by the group. By reviewing outliers, discrepancies and variations, the group should strive to reach consensus regarding the level of performance for the practices. During the second meeting, the group should also discuss the assessment process in order to continually refine the process based on considerations specific to the enterprise (feedback may include additional respondents, more training, other key information to consider, etc.). Finally, once scores are set, the third meeting provides an opportunity to evaluate the results and translate the assessment outputs into interpretations, insights and actionable projects to improve performance in critical areas. This final session should be more inclusive then the assessment, as it has impacts on all enterprise users. Involving a greater number of stakeholders will result in more ideas being brought to the table and greater commitment to action plan for improvement.

Introduction and Training. The last step in assessment planning is to conduct training for the both the facilitators and users. This training provides an opportunity for the facilitator to convey core assessment principles (lean concepts and practices), key terminology and information about the assessment to the stakeholders. During the training, it is important to convey the chosen boundaries and a transformation timeline, as this establishes a baseline for assessment ensuring that the scores provided by the respondents are based on consistent notions of the enterprise and the transformation plan. Overall, the introduction and training step helps reduce bias, prepare respondents for the assessment and ensures that the results are both valid and usable.

3.1.6. Phase Three: Perform Assessment

Now that respondents have been selected and trained, the actual assessment can take place based on the timeline established.

Conduct Individual Assessments. Each respondent will must read and score the 54 practices in LESAT, measuring both the current state of performance and setting a desired state based on the constraints of a predetermined transformation timeline (that is, all practices should not be given the highest possible score for the desired state). Respondents should also provide comments for as many of the practices as possible. LESAT has space delineated for both evidence (data or insights that support the chosen current state) and opportunities (projects or ideas that will help in achieving the desired state). By providing comments, the evaluation and planning process will be more efficient and will reflect a diverse range of ideas for improvement.

The scoring of practices can be done in a common space, or respondents can take the assessment document and complete the scoring independently. Having a common space will allow the facilitator to be present and provide clarification in a way that will benefit all respondents. In addition, some respondents may find it beneficial to complete the assessment with their direct reports in order to make it more collaborative and to provide scores more reflective of their department/unit.

The medium for scoring may vary based on the enterprise. LESAT can be done either as a paper assessment, or using an Excel worksheet, or online via an automated tool. Using electronic tools can increase efficiency for analysis and collation, but may be harder to coordinate and consolidate for a group session.

Respondents should be given basic instructions about the LESAT practices during the actual assessment session (these instructions can also be covered in the training, but it is important to reinforce them prior to the assessment process). Key points of the instructions include:

- Respondents should read descriptions of each level of performance in ascending order (from left to right), and select only select a level if the enterprise meets the description of proceeding levels. If they feel the enterprise is between levels, they should select the lower of the two.
- Respondents should score each practice based on the scope of the enterprise, and the desired score should reflect what they consider realistically achievable given the transformation timeline.
- Respondents should try to score all practices, and should only skip a practice if they do not understand it, they are not aware of performance, or they feel it is not relevant to the enterprise.

These guidelines help ensure consistency in the responses, enabling better comparison of scores across respondents.

Collect and Process Results. Once the respondents have provided scores and comments, the facilitator should collect and collate all data. The respondent should prepare relevant graphs and highlights practices of interest. Because of the richness of LESAT data, there are numerous methods for parsing and presenting the assessment results. A subsequent subsection (3.2.1) in this chapter is devoted to the different recommended methods for analyzing and presenting LESAT results.

Discuss and Analyze Results. Now that the facilitator has prepared summaries of initial results, the facilitator can present this information to the respondents for discussion. This provides an important opportunity for gaining consensus, understanding variations in scores and determining final assessment results. The group should look for unusual practices (those whose scores substantially differ from other practices) or practices where there is a significant variation

in the scores provided (a lack of consensus). By discussing these practices and their implications, the group can begin to reach a consensus for unusual scores and can begin to identify opportunities for enterprise transformation. During this discussion, respondents should work to clarify what is meant by each practice as well, correcting errors that may have resulted from misunderstandings. Throughout the process, the facilitator should strive to create an environment that fosters honest and constructive discussion between all the respondents.

3.1.7. Phase Four: Evaluate Assessment Results and Process

The previous phase resulted in the enterprise reaching consensus and obtaining preliminary insights based on the scores. The next phase takes the specific outputs of LESAT and interprets them in the face of the unique context of the enterprise in order to reach insights that will drive strategic planning and the transformation process.

Evaluate Assessment Results. In order to aid with the evaluation of results, the assessment users should establish preliminary decision criteria upfront that maps practices into different categories. Having predefined decision criteria helps in categorizing practices (reducing the complexity of interpreting 54 different practices) and will help standardize the evaluation process, avoiding unnecessary disputes that might accompany a more subjective interpertation of practices. Decision criteria combine the assessment results (like current, desired, gap and variance) into broader insights, and several potential decision criteria are introduced in the subsequent section (3.2.2) on assessment interpretation.

Once decision criteria have been applied, respondents can then discuss the preliminary classification of practices and work to refine the results into strategic focus and key actionable areas for the transformation plan.

Evaluate Assessment Process. During this phase, respondents and assessment users should also discuss the assessment process itself in order to identify areas for improvement and set a reassessment timeline. The discussion should focus on a number of key areas regarding the assessment process, including effectiveness (were the desired objectives met?), timeliness (were the insights available when needed?), efficiency (did the process go smoothly?), choice of respondents (were a variety of enterprise perspectives represented?), accuracy and validity (did the process reinforce usable results?), and obstacles (what factors hindered the assessment process?). By reviewing these topics, the respondents and users can provide key feedback to the facilitator and draft a revised process for subsequent assessments. This iterative nature will allow assessments to become increasingly efficient, relevant and timely.

3.1.8. Phase Five: Develop Action Plan and Prioritize Resources

The final phase of the assessment process is to put the results and evaluation into actionable improvement plans that are supported by the necessary resources. Based on the areas for improvement identified in the previous phases of the process, a distinct subset should be selected based on existing capabilities and on their importance for the enterprise strategic objectives. Based on the processes identified in the guiding transformation roadmap, the enterprise leaders should prioritize these core areas for improvement and allocate the necessary funding and support to drive improvement efforts.

This is one of the most important phases in the assessment process. This provides the closed-loop integrated control necessary to tie assessment outputs into actionable projects and identifiable improvement initiatives (Hallam, 2003). Nonetheless, this step of the assessment process will not be addressed in this thesis. Instead, the methodology for creating an abridged

assessment and associated abridged assessment process in the subsequent chapters primarily affects the second, third and fourth phase of the assessment process.

3.2. Insights from a Single Assessment

In this section, we use literature and case studies to enumerate methods for analyzing and interpreting assessment results in order to drive the transformation process. We start by looking at basic methods of summarizing results: averaging current performance scores for an overview of current strengths, calculating the gaps between current and desired performance to prioritize transformation tasks and using outliers to provide insights into problems and opportunities by leveraging the multiple perspectives inherent in the self-assessment process. Another important result of the self-assessment process is the variance between each respondent's current performance scores, which provides insights regarding leadership cohesion and agreement.

Next, we look at ways to extract insights and categorize practices based on relative performance differences. By grouping practices into key categories (like those in a SWOT – Strengths, Weaknesses, Opportunities and Threats – analysis), the assessment results can be used to guide and assist the transformation process. We find that the scores and results help highlight strategic strengths and areas for improvement, but are only provide a portion of the input needed for understanding performance and driving transformation. Analysis must also rely on the individual comments and broader contextual factors that influence both the enterprise and the assessment.

We also look at some of the factors that influence the reliability and validity of such assessments (both specific to the actual assessment and to the organization interpreting the results). Factors such as participant selection, assessment process and inter-practice/inter-

assessment comparisons can affect results, and should be accounted for during the analysis process.

Finally, we examine some of the crucial first- and second-order benefits and outputs from the analysis of ongoing assessments. Regular assessments support strategic planning, enable alignment between day-to-day decision-making and strategic direction, and increase awareness and knowledge of key concepts. These benefits reinforce the value of enterprise assessment, and the need to carefully distill insights from assessment results.

3.2.1. Analysis and Presentation of Primary LESAT Outputs

Much of the methods for analyzing LESAT results are based on the Facilitator's Guide (Abdimomunova, 2010) and the work of Hallam (both his dissertation and his presentation on interpreting and acting on LESAT results) (Hallam, 2003; 2004). There are several basic parameters that can be used to assess LESAT results, and these parameters can then be combined into graphs and histograms to communicate key insights from the assessment process. Throughout this section, we include illustrative graphs and data from a range of enterprises and have been normalized to ensure that there are no identifiable details.

During the analysis process, it is important to reinforce the relative nature of interpretation. Because of the contextual and environmental differences that influence responses in a self-assessment exercise, it is important to look at data in a relative manner as opposed to an absolute manner. For example, many of the interpretation strategies look at high or low scoring practices. Rather than looking at the absolute score, the differentiation "high" and "low" should be based on a relative comparison of all internal scores. As a result, what constitutes a high or low score will differ between organizations and between each individual assessment period. In order to accurately and consistently identify relative differences, responses can be broke into

different qualities when comparing scores from a common assessment, while comparing multiple assessments requires more complex normalization. Potential normalization strategies can be drawn from research in collaborative filtering that looks at equating consumer preferences across different rating models and criteria (Lansing & Si, 2004).

Analyzing Current and Desired State Scores. One of the best starting points for analyzing results is to look at average current and desired performance scores across respondents. For current performance, the average provides some insights into current lean adoption and current enterprise maturity. In this way, the average score serves as an overall health status for each practice within the enterprise. Low current scores represent practices that have not been adopted yet and require more focus, while high current scores represent organizational strengths and successes. This knowledge is crucial for designing a transformation plan, where the organization must understand its current performance and current strengths in order to map out a transformation plan that builds on these strengths and improves the weaknesses.

Average desired scores provide feedback on the importance of different practices to the organization. The desired scores can be seen as reflecting two things: the priorities and values of the organization, as well as the weaknesses. High average desired scores represent practices that are extremely important to the organization, and have substantial benefits. Low average desired scores might represent low priority practices, or they may represent overlooked practices that have not yet been capitalized upon; as a result, the respondents do not realize the potential benefits that can be derived from increasing performance for these practices.

Averages are used for interpreting scores in order to represent the aggregate opinion of all respondents. Median scores are less susceptible to outlying scores, but are also less

representative of the opinions of all respondents. As a result, the average of all respondents provides a better mechanism for a holistic performance assessment. Later, we discuss some potential interpretations of outlying scores.

Analyzing Gaps. By subtracting the current score from the desired score, a gap can be calculated for each practice. Gap scores are extremely useful in understanding current high-priority practices to target in the transformation plan.

Large average gaps represent practices with low current scores and high desired scores, meaning there is room for improvement and respondents generally see substantial value in achieving improvement (hence the high desired score). Small average gaps can be representative of two things. One, if both the current and desired score are high, it suggests that the organization has achieved a high-value goal and should ensure that performance remains high through ongoing observation and assessment (ideally, organizations should develop leading indicators or metrics to ensure that performance on such practices remains high). Alternatively, if both the current and desired score are low (an unusual outcome), the small average gaps suggest that a practice is not prioritized or valued within the organization and that this is reflected in current performance. By discussing the practice, the organization can determine if the practice really is low priority (that is, it offers no significant gains). Or, if there are potential gains, the practice may actually represent unrealized potential, where the value and benefits of increasing current performance have not been communicated or prioritized by the organization's leadership and management.

By creating a Pareto graph of the individual practices according to gaps, organizations can identify those practices which respondents see as being significant opportunities for growth (Figure 2). This can aid in prioritizing performance-improving measures based on the level of

support and consensus (of course, process-improvement decisions should also reflect other considerations such as cost and schedule constraints). If a practice has a large gap, respondents have identified it as an area where improvements will yield substantial benefits. The consensus among respondents suggests an area with a high perceived return on investment and, as a result, a commitment from respondents to realize change, both of which warrant moving forward with improvement efforts targeted at the practice in question. On the other hand, those practices with low gaps tend to have a lower level of commitment and a lower value proposition.

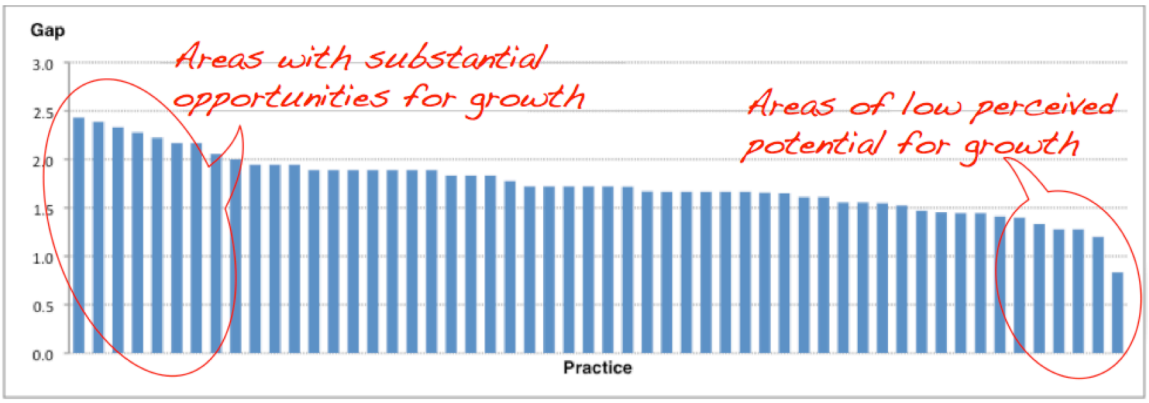


Figure 2 – Gap analysis graph

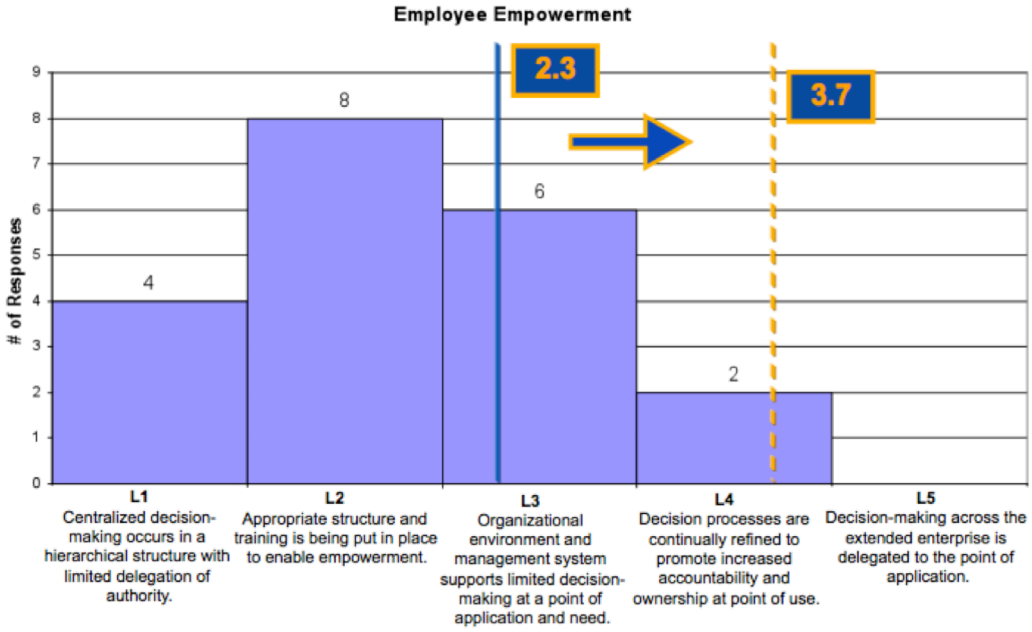
Tallies and Ranges. A useful method of combining results from all respondents is to create a table depicting both the range and tally of responses for each practice. This analysis provides a complement to the single average score, by providing an overview of how responses varied. The range provides an idea of the overall fluctuation in responses (as with the average, the range is used because it is provides a very inclusive view of all responses including outliers, unlike the standard deviation). In addition, showing a tally of responses can be indicative in understanding the distribution of scores provided to each practice. Table 3 shows an excerpt of a large table depicting ranges and tallies for each practice. The columns labeled “L1”, “L2”, etc. include the total number of respondents who scored the practice at that current level.

Table 3 – Excerpt from LESAT summary table

LEAN PRACTICE	CURRENT STATE								D	G
	Mean	Var.	Range	L1	L2	L3	L4	L5		
I.A.1. Integration of lean in strategic planning process	2.5	0.4	2.0	0	11	8	1	0	4.1	1.6
I.A.3. Leveraging the extended enterprise	2.4	0.6	2.0	3	7	10	0	0	3.8	1.4
I.B.2. Senior management commitment	1.4	0.6	3.0	14	5	0	1	0	3.4	2

Note: *"Mean" – Average current state score*
"Var." – Variance of scores
"Range" – Range of scores, i.e. the difference between the highest and the lowest score
"L1", "L2", "L3", "L4" and "L5" – Capability levels 1, 2, 3, 4 and 5 respectively
"D" – Average desired state score
"G" – Gap between the average desired and the average current state score

One method for summarizing and communicating the key insights of this data is to present it in a histogram of current scores for key practices. A histogram (Figure 3) helps visualize the distribution of current scores, and by overlaying both the average current score and average desired score, it highlights the gap that needs to be addressed by the transformation plan. Such presentation and analysis of information can help reveal information about the diversity of perspectives within the respondent body.



Note: The blue vertical line is the average current score, while the yellow, dashed vertical line is the average desired score.

Figure 3 – Histogram of current scores for a LESAT practice

Variance. The above analysis, looking at both ranges and response distributions, is all targeted at understanding one key idea: variance in responses. In many ways, variance is an extremely important output of the LESAT process, possibly more important than the practice scores. Each respondent provides his or her scores for the 54 practices, which provides an understanding of the current state or organizational maturity. But variance is an endogenous indicator that results from combining the scores from each respondent and represents agreement and communication across organizational departments or structures. Low variance suggests a high level of agreement and cohesion across the group of respondents, while high variance suggests the opposite. Of course, this interpretation of variance is dependent on having a broad sampling of respondents representing different divisions and aspects of the enterprise, which is an important phase of the assessment process (Abdimomunova, Perkins, Valerdi, Shields, & Nightingale, 2011).

Variance can also be looked at on a departmental level (for example, variance between departments may be high). For example, if IT enabling processes are only rated highly by IT personnel, this suggests that their values are not aligned across the organization. Although all respondents are instructed to assess the enterprise as a whole, it is inevitable that the scores from each respondent can vary due to their perspective and individual insights into enterprise performance. In addition, areas of high variance can suggest a lack of lean knowledge, resulting in different interpretation of the scoring levels or criteria.

As with gaps, creating a graph depicting the range of variance can be extremely useful for directing focus (Figure 4). Practices with high levels of variance require further discussion, as well as provide potential areas for training and education, to increase knowledge and understanding of the practice. On the other hand, practices with low variance suggest stronger cohesion, communication and cooperation across organizational boundaries. Of course, low variance combined with low scores means that there may be cohesion and common understanding, yet there is still much room for improvement.

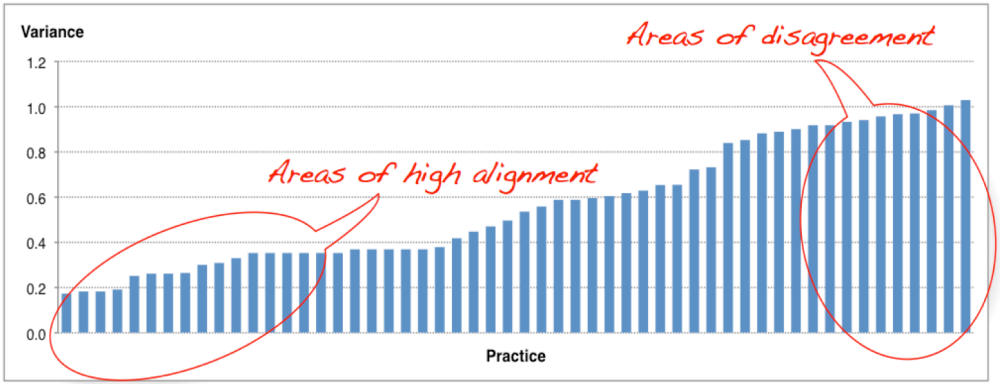


Figure 4 – Ranking of LESAT practices by variance

Summary of Initial Analysis. Each of the interpretation methods discussed thus far looks at a single output of LESAT, and they yield useful interpretations and insights as

summarized in Table 4. These initial insights are useful, but much of the power of LESAT comes from combining these different indicators. Later sections show tools for combining indicators or delving into further levels of analysis.

Table 4 – LESAT indicators

	Low	**High**
Current State Score	Low current performance or enterprise maturity	High current performance or enterprise maturity
Desired State Score	Low priority or low realized benefits (could be unrealized benefits)	High priority and high potential benefits
Gap	Close to achieving desired state, could be under realizing value of practice or could be a low priority practice	Area with low current performance and high realized gains making it a high priority for improvement
Range and Variance	High agreement on the level of performance and understanding of the practice at hand	Low agreement and understanding of the practice, or different views of performance across respondents/divisions/departments

3.2.2. Looking Deeper: Methods for Combining Multiple Indicators

The below techniques and strategies represent the methods for more detailed analysis of scores as set out in the Facilitator's Guide (MIT, 2001).

Averaging Across Parts or Sections. Although each practice deserves analysis and review on an individual level, there is also merit in averaging across the three structures and/or the twelve subsections used to group practices. This is particularly useful in using LESAT as a leading indicator for improving organizational performance. Evidence has demonstrated that improving performance in lean enterprise transformation and leadership (Section I) leads to improving performance in both lifecycle processes (Section II) and enabling infrastructure processes (Section III). Improving performance in enabling infrastructure processes (Section III)

ties with further improvements in lifecycle processes (Section II) (Hallam, 2003). These results both provide added weight and evidence for the value of LESAT, but also help illustrate the interconnected nature of the three classes of processes and the residual benefits from achieving improvements in specific areas. The interrelationship between section scores can aid in understanding transformation outcomes and can be used in the prioritization of future transformation efforts, as the connected nature reinforces the need for a groundwork of leadership and enterprise thinking (Section I) as well as supporting processes (Section III) before lifecycle improvements (Section II) are realized. Having this understanding can guide the direction of ongoing transformation efforts, and reinforces the holistic nature of enterprise performance that LESAT seeks to assess.

Although section and subsection scores are not sufficient for analysis, they do provide an overall picture and useful tool for highlighting certain focus areas (broader than a single practice).

Outliers. Invariably, certain practices or certain respondents may introduce outlying scores to the assessment process. These can be valuable to interpreting the scores, especially when evaluating with regards to the respondent's position. Certain contextual factors may contribute to the outlying nature of their responses, such as the length of their employment (new employees vs. long-time employees), their leadership position, managerial influence or their relation to the practice at hand. These contextual factors must be considered when looking at outlying responses. If these alone do not account for the outlier, further discussion with the respondent may be valuable for understanding the outlier, as it may be the result of a misunderstanding or misinterpretation of the practices, or it may be a unique insight about the

enterprise. Only those outliers that are due to mistakes should be eliminated, as the other ones provide important evidence of external or contextual factors influencing the outlying response.

Comparing Respondents' Positions or Roles with Scores. Along the lines of the discussion of outliers, comparing a respondent's scores with their position in the organizational hierarchy can be useful for understanding variance in certain questions. Although these may highlight sensitive issues of disagreement, they can be useful in understanding certain obstacles in the transformation plan. An example that often comes up during the LESAT process is leadership commitment, where high-level leadership rates their commitment as high, while other respondents report less commitment and support from top-level leadership. Such a division reveals differences between perspectives and reality, which can otherwise hinder the transformation plan. Only by addressing these variations and ensuring all personnel are in agreement can an adequate and relevant transformation plan be developed. For this reason, analyzing score variation against position in the organizational hierarchy is an important step in interpreting LESAT results.

Of course, it is also important to preserve a sufficient level of anonymity to avoid managerial influence (which may cause respondents tailor their scores for their superiors, in order to appear more optimistic). When presenting such data, the individual responders should generally not be discernable. How this should be handled depends on the openness of the organization.

Using Comments on Evidence and Opportunities. The LESAT provides broad focus areas for enterprise transformation in terms of areas with low current performance and a high level of common-valuation among respondents (i.e., high desired performance). Yet these insights provide only a starting point for the broader transformation path, as the practices in

LESAT are often broad and can be approached through several avenues. Encouraging respondents to provide comments, then analyzing the identified evidence and opportunities for each practice can provide an important starting ground for designing improvement projects to achieve realistic improvements in existing business activities. Ideally, comments related to evidence for each practice would help to determine high-performing departments or areas of the business. These departments can be used as a source for documenting best practices for implementing the specific practice across the broader enterprise. Likewise, comments related to opportunities for each practice will help target initial improvement projects related to the practice targeting those areas or departments that might most benefit from increased the performance.

3.2.3. Interpreting the Results

Once the initial analysis and summary work is completed, it will be possible to identify and classify different practices in relationship to the enterprise ("highest performing", "areas for improvement", etc). The next step is to delve further and look at the interrelationships between the practices and the identified classifications. This more in-depth interpretation can provide an important precursor for transformation planning, as it allows for a holistic view of the distinct practices. In addition, these interpretation methods seek to clarify the results into key graphics that provide a summary of essential results enabling the dissemination and communication of LESAT results back to enterprise leaders and staff.

During data collection and discussions with LESAT users, we have learned about novel ways for interpreting results. Below we describe some strategies that can assist users in identifying priority areas for improvements to be addressed in the transformation process.

Mapping Current State Score vs. Gap. One of the first ways to combine the different indicators is to create a simple scatter plot to highlight areas for improvement (Figure 5). The

horizontal axis represents the current score of each practice, while the vertical axis represents the gap. The illustrative graph below has data points for all 54 practices, but alternatively, the scores for each part of the LESAT can be grouped together based on the LESAT structure (see Appendix A: LESAT Structure) to create fifteen general focus areas (e.g., "Enterprise Strategic Planning", "Business Acquisition and Program Management", etc.).

The scatter plot is a useful tool for prioritizing change. One novel method for interpreting LESAT results involves mapping the practices into the categories of a SWOT (Strengths, Weaknesses, Opportunities and Threats) analysis. This method allows using assessment scores to identify current strengths and weaknesses in the organization, as well as the opportunities and threats that it is facing (Figure 5). Practices with the highest current state performance represent existing *Strengths* of the organization, while those with the lowest current performance can be interpreted as *Weaknesses*.

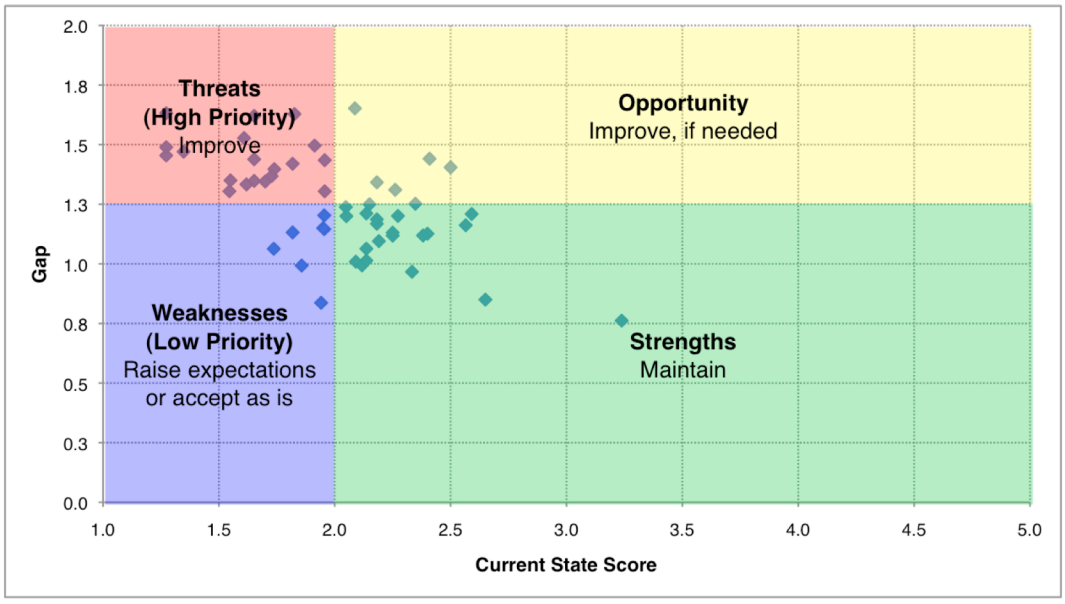


Figure 5 – SWOT analysis superimposed on LESAT practice scatterplot

High priority practices are those found in the top left of the graph, as they have both low current scores and high gaps. They can pose a *Threat* for the organization, if not dealt with. The low current state score indicates that the organization was either unaware of the practice or has not paid attention to it so far. It suggests substantial room for improvement, and hence substantial opportunities for gains. However, the high desired state score might suggest that there is a realization about the important role and the potential gains of the practice in the transformation process. By focusing transformation exercises on such practices, the organization has substantial opportunities for improvement with a strong level of buy-in. At the same time, achieving the high desired state may be a challenge due to the high gap. It may require additional resources and time as well as overcoming resistance within the organization.

A combination of high current performance and high gap indicates that the organization has high aspirations with regard to the assessed practice, while the current state suggests that certain progress is being achieved in the enterprise transformation. Such practice presents an *Opportunity* for improvements, as the organization has already started to work on them and is committed to achieve a higher level. Due to this, the improvements on these practices may be easier to achieve, benefiting from past improvement efforts and strong commitment. On the other hand, as performance increases, organizations will face a plateau effect, where further improvements become more difficult (Godard, 2004).

As with all interpretation strategies, the definition of a high and a low score depends on the particular scoring results within the organization. The terms *high* and *low* should be interpreted in relative terms vis-à-vis other scores across the full set of practices.

Mapping Gaps vs. Variance. Another novel and beneficial tool for interpreting LESAT results is the mapping of gaps to variances (Montoya, Schofield, Chowdhury, & Lehman, 2009).

As in the previous section, start off with a simple scatter plot, where the horizontal axis represents variance of the current score of each practice, while the vertical axis represents the gap (Figure 6).

The scatter plot provides a tool for categorizing practices based on a combination of both their variance and their gaps, in order to address the most beneficial practices first (Figure 6). Practices with high gaps and low variance are considered *Fertile Ground*, as there is a high level of consensus regarding the need for improvement. If there is a high gap and high variance, the practice needs further *Open Discussion* or may benefit from additional training in order to create better alignment between respondents. If the gap is low and there is a high level of variance, the practice can be categorized as a *Low Alignment*. Such practices are low priority for transformation, but are good areas to increase knowledge and understanding. Finally, practices with both a low gap and a low variance are *Low Priority* in general. As discussed before, these practices may also represent unrealized gains, and further attention should be paid to such practices to identify whether there may be worthwhile gains in improvement.

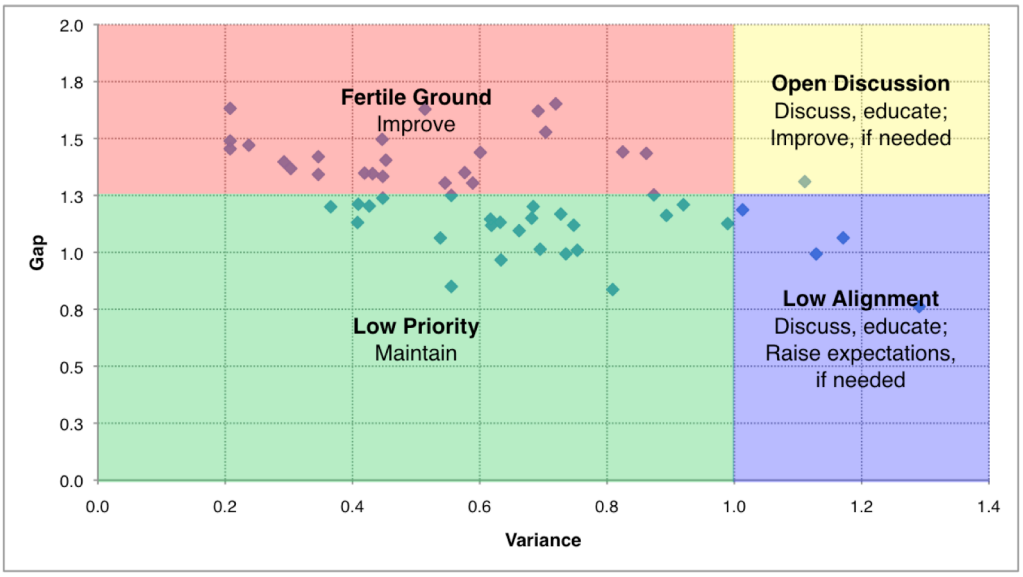


Figure 6 – Variance vs. gap scatter plot of LESAT practices with interpretations

Three-Dimensional Analysis. Another interesting method for analyzing LESAT results combines the above two techniques. It was developed by lean practitioners at an industry partner. This method was used for analysis of the results obtained during the use of LESAT to assess functional units, and was shared with the Lean Advancement Initiative during a knowledge exchange event in 2009.

LESAT scores were analyzed across three dimensions: current state performance, variance and gap (see Table 5). As demonstrated earlier in mapping of results onto SWOT analysis, the practices with high and low current state performance represent strong and weak areas, respectively. Relatively high performing practices need to be maintained at their current level, unless some further improvements are deemed necessary. On the other hand, the organization needs to take action with regard to low performance practices in order to achieve improvements.

The second dimension, variance, represents the level of agreement within the organization on each practice. The practices with relatively low variance of scores indicated strong agreement among respondents with regard to the current state performance. The practices with high variance may suggest lack of agreement among respondents and lack of awareness about the approaches being taken by the organization. For such practices, additional education may be required.

The third dimension, gap, represents opportunities. In case of low gap, the organization may be presented with an opportunity to raise expectations with regard to the practice and accept the current state as it. If so, no action is required to be taken. If the gap is high, however, improvement actions need to be taken to close the gap.

Upon the analysis, some practices may emerge as fitting criteria in more than one dimension. The analysis can be used to prioritize transformation efforts; practices with low current performance and high gap must be addressed initially. Practices with high current state, high variance and low gap may be addressed through education. No actions are required to be taken for practices with high current performance and low gap.

Table 5 – Three-dimensional analysis of LESAT scores

Metric	Level	Example Score Range*	Interpretation	Recommendation
Current State	High	≥ 2.5	Strongest areas	Maintain or improve
	Low	≤ 1.5	Weakest areas	Improve
Variance	High	≥ 1.0	More disagreement. Need for training/education	Education needed
	Low	≤ 0.5	Strong agreement	No education needed
Gap	High	≥ 1.5	Opportunity to close the gap through lean improvement	Take action
	Low	≤ 1.25	Opportunity to raise expectations or accept as is	No action

** Score breakdowns should be adjusted relative to distribution of scores.*

3.2.4. Common Mistakes

The three methods above discuss novel and beneficial tools for interpreting LESAT results, but during data collection and feedback opportunities, we have also heard about several ineffective ways of interpreting or using LESAT results that undermine the value of the assessment process. We will now look at a few of these common mistakes.

Losing Fidelity. The first involves losing fidelity by oversimplifying scores. LESAT represents a multidimensional assessment of broad lean adoption. Each practice is independent. Although they can be grouped into parts and sections, it is detrimental to average all scores

together and simply look at LESAT results on a section-by-section basis. Likewise, averaging all practices into one single score undermines much of the value behind LESAT. Having a singular score is appealing from a management perspective, as it provides the a clear indicator for improvement, but doing so undermines the granularity and fidelity provided by the multitude of practices included in LESAT. For example, if all scores are averaged together, improvements in one set of practices may be masked by declines in another set of practices. Without this level of understanding, the organization cannot adjust its transformation plan and introduce corrective action, nor can it see its successes.

In addition, comparing and averaging scores across practices faces some fundamental constraints. Although the five levels were designed to be consistent for all practices, this is not always the case. The assessment is prone to scale inconsistencies, which are only exaggerated when LESAT viewed from the perspective of a respondent or an organization (as their interpretation of the scale may easily differ from the interpretation used by the authors of LESAT). As a result, comparing or averaging scores for different practices can be like comparing inches with pounds, due to the fact that the perceived scale may vary between practices.

As mentioned earlier, there are benefits from averaging scores as it sheds light on the interconnections between sections. Yet section averages should be used cautiously, in order to not detract from the in-depth, per-practice analysis.

Not Interpreting Results. One of the biggest mistakes that can be made is to perform the assessment, but not interpret the results. When used this way, it forms an open loop assessment, where results of the assessment process have little or no direct effect on the transformation plan (Hallam, 2003). This can be manifested in a few possible scenarios: where results are

superficially analyzed (averaging all scores together), or where results are analyzed but not distributed to key parties or acted upon. Although the organization will experience minor improvements in lean enterprise knowledge and vocabulary, there will be few gains, since the results are not interpreted and do not guide ongoing transformation or continuous improvement projects.

Along these lines, some organizations use LESAT in parallel with other improvement plans. These organizations gain some of the benefits from interpreting LESAT, but must put added effort to coordinate contrasting foci and allocating limited resources among the improvement plans.

Not Accounting for External Context. Important for interpreting LESAT results is a strong understanding of the organizational context and environmental factors that may have a bearing on the results. There is a substantial number of factors or events that may affect the results of the assessment process, and accounting for these is important to understanding results. Example factors include: changing leadership, recent shortcomings or major successes, increasing lean knowledge (and hence changing interpretation of questions), bias related to the person facilitating the assessment, etc. There are a multitude of factors that may have a bearing on the results. When analysis of LESAT scores fails to account for such factors, the benefits and value of assessment are weakened.

3.2.5. Summary

We have reviewed a number of different strategies and tools for interpreting and analyzing the results of a LESAT assessment. The interpretation strategies and tools that will work for a specific enterprise vary, depending on the organizational needs, the transformation framework being implemented and the facilitator's familiarity with LESAT. But the tools and

strategies listed above provide a good starting point for analyzing data and identifying preliminary take-away messages. Additional tools and analysis may be used based on the contextual and environment factors that influence the enterprise being assessed.

In addition, reviewing the comments on each respondent's survey (evidence and insights for each practice) can help guide the interpretation process, by providing a more complex response that can aid in interpreting the meaning of the scores.

The interpretation of assessment results and the insights gained from the assessment process provide crucial guidance for transformation efforts, by identifying high-priority improvement areas, organizational strengths and weakness and an understanding of needed areas for improvement. Of course, the assessment results cannot act as the sole input for the transformation process. Other factors play an important role in influencing the transformation process, such as the order of practices. As outlined in the Enterprise Transformation Roadmap (Nightingale, 2009; Nightingale & Mize, 2002), there is a distinct order to the transformation process where certain LESAT practices are prerequisite for subsequent improvement efforts. The interconnections between practices and the prescribed roadmap provide additional insights in how to translate assessment results into a clear transformation plan.

3.3. Benefits of Ongoing Assessment

The previous section discusses the insights and interpretations that result from a single use of LESAT, yet there are additional insights that result from ongoing use. LESAT was created with the intention of regular use and reassessment as part of the continual improvement journey. The benefits resulting from ongoing use of LESAT are presented in the "*First Order Benefits of Analysis*". We discuss additional benefits in the "*Second Order Benefits of Analysis*".

These benefits are especially relevant to abridged assessment, which are most advantageously used as a reassessment tool – that is, providing less resource intensive method for tracking assessment results on an ongoing basis. By making reassessment more appealing, enterprises are better able to capitalize on these additional benefits of assessment.

3.3.1. First-Order Benefits of Analysis

The frequency of reassessment can range depending on the organizational needs, but should be set in accordance with the transformation plan and objectives, as reassessment provides crucial a crucial status and health check on transformation progress. Higher frequency reassessment provides additional opportunities to monitor transformation progress and gain feedback on success, but also results in a higher cost in terms of time and resource devoted to assessment. Based on this trade off, many organizations opt to reassess annually or biennially.

Much of these ongoing insights depend on having a relatively consistent and regular set of respondents completing the LESAT. If there is substantial fluctuation in the respondents or the positions and perspectives represented in the group of respondents, then it will become more difficult to compare results over time, as fluctuations in scores may be the result of changes in the pool of respondents, rather than substantial performance changes.

Regular ongoing assessment provides a number of tangible, direct benefits to the enterprise by providing important insights into performance and organization trends.

Internal Benchmarking. One of the strong advantages of regular reassessment is a greater ability to benchmark the LESAT results. Because of the varying contextual factors that affect both the assessment process and the assessment results, the ability to have internal benchmarks is useful for understanding and interpreting the changes observed during reassessments and maximizing the benefits of LESAT. By performing an initial assessment at the

beginning of the transformation process, baseline scores can be established. These initial scores help in setting transformation priorities, but also provide a crucial benchmark for viewing and interpreting reassessments associated with the transformation plan.

As discussed before, research has established LESAT as a leading indicator for lean transformation, showing how improvements in lean transformation and leadership (Section I) and enabling infrastructure (Section III) are associated with improvements in lifecycle processes (Section II) (Hallam, 2003). By establishing a history of LESAT surveys, this trend can be used to interpret and respond to shifts in scores. For example, increasing scores in one section suggest scores will increase in other sections. By focusing transformation efforts on these sections, the increases can be maximized. Alternatively, if scores are found to be decreasing in one section or part of LESAT, this warrants further investigation and analysis, as well as corrective action, in order to avoid more widespread decreases in performance. Especially for the early reassessments, it is common to observe a decrease in scores that is due to increased understanding of the full extent of the practice by respondents.

Overall, regular reassessment and internal benchmarking can be very useful for observing trends and gaining an understanding of LESAT tailored to the localized interpretations and uses of the assessment. The ability to compare current performance with past performance is invaluable in understanding the meaning of LESAT specific to the enterprise being assessed.

Benchmarking Across Enterprises. In addition to internal benchmarking, regular LESAT use across an extended enterprise offers opportunities for benchmarking across enterprises and setting widespread improvements or targets. This benchmarking across enterprises has been seen in many aerospace companies, with geographically diversified management and manufacturing operations. In such situations, the LESAT assessment is

conducted at each enterprise, to look at the performance at that geographic site. Then, results and transformation plans can be compared across the extended enterprise to understand common challenges and common opportunities that can be addressed by the top-level leadership in order to support the transformation plan. As reassessments are performed, shifts and improvements can be compared across enterprises in order to encourage improvement and detect outliers.

Tracking Progress in Transformation Journey. The primary motivation for ongoing reassessment is to track progress in the transformation journey. This was discussed previously with regards to the internal benchmarking, but it is an important point and deserves elaboration. Regular reassessment is crucial to gaining insights into the success and progress of an ongoing transformation journey. Depending on the broader transformation framework, this information can be tied into refining transformation plans, shifting resource or focus allocation, or other adjustments in order to increase the success and outcomes of the transformation plan. This can be achieved by simply comparing the current performance as scored in the reassessments with the targets and goals of the original transformation plan and the original assessment's desired performance scores. Those practices that have approached the goal are areas of success, while lagging practices (those that have not experienced a performance increase since the original assessment, despite related projects in the transformation plan) require more attention and potentially more resources in order to foster the needed change.

Tracking Personnel Cohesion and Communication Scores. Regular reassessment can also offer insights into increasing personnel cohesion and team communication, by observing trends or shifts in the variance among respondents. Research has shown that variance naturally increases as practice maturity increases, due in part to a broader range of possible performance levels (i.e., low maturity practices generally have lower variance because the performance

clearly matches a specific level, but as maturity increases, there becomes greater variation and subjectivity in classifying the performance level) (Hallam, 2003). Once this trend of increasing variance with increasing maturity has been accounted for, the variance during reassessment can be used as a judge of shifts in team cohesion and communication levels.

Tracking Lean Knowledge. By looking at respondents' comments and scoring trends over time, regular reassessment also offers useful insights into changing lean knowledge and understanding throughout the enterprise. As individuals become more familiar with lean practices and concepts, this will be reflected in their scores and comments (for example, the comments collected for evidence and opportunities may provide more concrete examples of opportunities and evidence based on the respondents better understanding of lean business practices). As for scores, it is hard to say whether increasing lean knowledge will shift scores in a particular way: increased knowledge may result in lower scores, because respondents better understand the practice they are scoring, or increased knowledge may result in higher scores, because individuals see more evidence and more payoff associated with the lean transformation process. These sort of across-the-board score shifts, as well as the comments with evidence and opportunities for each practice, are useful for tracking increases in lean knowledge and understanding.

3.3.2. Second-Order Benefits of Analysis

Thus far, we have discussed a variety of useful and powerful interpretations and insights that can be gained through performing LESAT as part of an ongoing transformation plan. These benefits represent some of the first-order, tangible benefits in terms of immediate gains and knowledge provided by the LESAT process. But conducting and analyzing the LESAT assessment regularly will also yield additional, second-order benefits throughout the organization

(Miller, 2006). These second-order benefits are of tremendous value to the enterprise and provide a compelling incentive for doing regular and in-depth analysis of LESAT results.

Assessment Feedback to Personnel. By sharing the analysis with the respondents, the respondents are able to see the realizations and output of the process. This feedback will provide respondents the feedback that their knowledge and insights are of value and are used in order to understand the best course of action for the enterprise. This feedback has residual benefits in motivating greater involvement in future assessment processes, as well as a greater appreciation for the value of assessment.

Driving Enterprise Behavior. By providing feedback to personnel, and by sharing the assessment analysis with organizational leadership and management, future enterprise behavior will be driven in part by the assessment results. As knowledge and awareness associated with the organization's current state increase, individuals will be more likely to take actions in accordance with the needs and goals of the enterprise. Overall, sharing analysis of enterprise performance will increase the commitment of personnel to the transformation plan and to improving the organization, because people better understand the motivation and need for change. As a result, enterprise behavior will be aligned with organizational needs and with the goals of the transformation plan.

These second order benefits from assessment are confirmed by research showing that behavior is predominantly aligned with what is measured (Hauser & Katz, 1998). But due to the complexity of the practices measured in LESAT and the self-assessment nature, the true behavioral benefits will only be realized once the results are analyzed and shared with personnel. This analysis will help provide key insights about focus areas and needs within the organization.

Enable Better Decision Making. The knowledge collected during the assessment process and shared back with leadership and management will enable better decision-making. When making decisions related to internal programs, training, improvement projects or organizational structure, decision makers will be equipped with the knowledge and insights that result from the LESAT assessment. This data can be used to direct decision-making and focus decisions on those actions that will be most beneficial to the enterprise as a whole. In addition, the LESAT will provide the crucial justification behind those decisions, helping ensure consensus and support for decisions regarding new initiatives or new projects.

Guide Transformation. In addition to driving general enterprise behavior and decisions, regular assessment will help guide transformation decisions and planning. By providing feedback and insights on both the current performance and past trends, decision-makers will be better equipped to design a transformation plan that meets the needs of the organization and will offer the highest benefit based on the values defined by all respondents during the assessment process. The ability of LESAT analysis to aggregate both an understanding of current performance and an evaluation of high priority improvements is invaluable to transformation planning.

3.4. Threats to Assessment Insights Validity

The LESAT is a powerful and useful tool in understanding an organization's performance, and helps direct transformation activities towards specific, high-value or high-priority practices. But as with any assessment tool, there are threats to validity that must be accounted for in the analysis process. Such threats may be internal to the organization, depending on how the organization impacts the assessment process and analysis. This impact may indirectly determine whether the assessment is going to benefit the organization or turn out to be a waste of resources. Many of these organization-specific threats to validity can be overcome through

careful planning or during the actual assessment process, by providing clear instructions and creating the trusting, safe environment needed to prevent bias (Abdimomunova, Perkins, Valerdi, Shields, & Nightingale, 2011).

Other threats to validity lie outside of the organizational boundaries and are attributable to LESAT as a measurement tool. In this section we briefly discuss two potential tool-specific threats to validity. It is important to be cognizant of limitations to the original assessment when creating an abridged version, in order to avoiding compounding existing limitations or threats to validity.

3.4.1. Organization-Specific Threats

Bias. As with many organizational feedback mechanisms, LESAT responses are vulnerable to biases and managerial influence. Results may be impacted by fear of managerial reprisal, a lack of anonymity, self-reporting bias (Donaldson & Grant-Vallone, 2002), confirmation bias (Nickerson, 1998) or system-justification bias (Jost, Banaji, & Nosek, 2004). This can result from overly optimistic scores. Although biases cannot be completely eliminated, steps can be taken to avoid and discourage them. By maintaining a level of anonymity and by fostering a safe, open environment for two-way communication regarding organizational performance, respondents will be less likely to bias their results or be biased by one view point. Of course, this sort of environment depends on having a committed top leadership, who are willing and interested to hear both the positive and negative feedback regarding current performance and opportunities for improvement.

As mentioned before, anonymity is both beneficial and detrimental to the LESAT process. Although it encourages more honest feedback, it also hinders the analysis by not providing information regarding each respondent's position or department. By having an external

facilitator conduct the assessment and by limiting knowledge of respondents to this facilitator, a middle ground can be achieved where analysis will reflect the variety of perspectives, but anonymity will not be compromised (minimizing bias).

Choice of Respondents. The scoring is also affected by the choice of respondents in the LESAT assessment. LESAT requests respondents to assess not only the current state performance of the organization, but also the desired state. Assessment of the desired state can be more realistic, and thus more meaningful, if the respondents understand and are able to evaluate resources needed to achieve that desired state. This requires respondents to have enterprise-wide views and responsibilities.

Optimistic or Pessimistic Respondents. Occasionally, one respondent may vary from the larger group by being more optimistic (higher scores) or more pessimistic (lower scores) in their responses. Generally, such respondents do not have a substantial impact on the interpretation (since analysis focuses primarily on relative differences rather than absolute scores), but this must be carefully judged during the analysis period. In some situations, the optimism or pessimism reflected in their scores might be relevant to the position or the context (e.g., a respondent about to retire may be more pessimistic). In other situations, it may reflect more of the general demeanor or outlook of the respondent. The analysis process must carefully account for such trends, and adjust for them accordingly.

Varying Interpretations. Each practice and scoring level in LESAT is subject to a certain level of interpretation by the respondent. This variation in interpretations can undermine the analysis, and is especially prevalent in organizations that are just beginning the lean journey (in such cases, there is often insufficient training regarding lean or transformation terminology). When analyzing results from an organization that is still embarking on lean transformation, the

analysis must look carefully at variance in scores and determine whether that can be accounted for by a lack of lean knowledge.

The impact of varying interpretations will decrease with time, as training increases and management begins communicating a clear and consistent message regarding lean and its application within the enterprise. The best way to mitigate this is through regular reassessment, where scores and results have the tendency to converge over time, reflecting increasing consistency in the interpretation and understanding of questions as they relate to the enterprise. Experience shows that as over time participants become more familiar with the lean culture, their scoring of the organization might in fact go down, as their understanding of LESAT practices improves and not because of the changes in the organization.

3.4.2. Tool-Specific Threats

Self-Assessment and Inter-Rater Reliability. The very nature of LESAT as a self-assessment tool, although advantageous in terms of aggregating several perspectives and taking a holistic view of the enterprise, has disadvantages that may impair the results of the assessment. One of the best examples of this is the issue of inter-rater reliability, which has to do with the level of variation between raters and variation in interpretations (LeBreton & Senter, 2007). High levels of variation may suggest an internal problem, due to a lack of training or coordination of lean efforts, but it also might suggest a weakness on the part of the assessment to adequately phrase questions in a way that elicits consistent and common responses. Further investigation is required to look at issues of inter-rater reliability to confirm the reliability of the LESAT assessment.

Scale Inconsistencies. During data collection and opportunities to hear about practical uses of the LESAT tool, one potential issue that was identified was inconsistencies in the five-

level scale across practices. For example, moving from a level two to a level three might be much harder for one practice than another. Although this does not suggest any degradation in the results of LESAT, it simply suggests that the initial prioritization of tasks may need to also take into account the varying difficulty of implement different practices are moving up one level vs. another level.

3.5. Motivation for Abridged LESAT

The motivation for an abridged assessment tool, and specifically an abridged version of LESAT, arose when reviewing case studies from organizations that only used LESAT once or twice. Although there are substantial insights to be gained from interpreting the scores from a single assessment (as covered in section 3.2), many ongoing benefits are not realized until assessment becomes embedded in the strategic planning and transformation cycles. Research has shown that using LESAT as the sole assessment tool in an integrated closed-loop control process generates the most substantial gains in performance for LESAT practices (Hallam, 2003). "Integrated closed-loop control process" involves using the outputs of the assessment tool to directly allocate resources and drive the improvement plans, as illustrated in Figure 7.

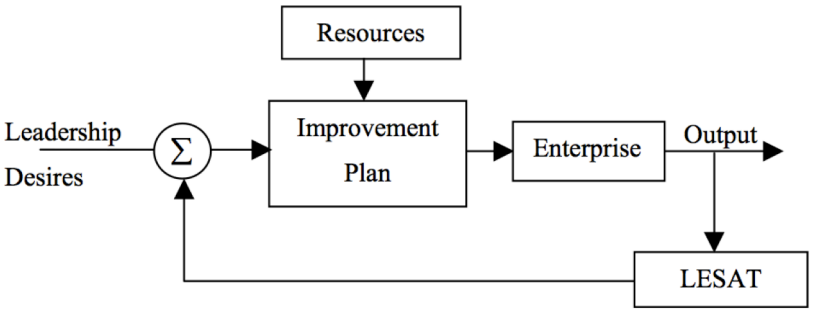


Figure 7 – Integrated closed-loop control (from Hallam, 2003)

In order for ongoing assessment and integrated closed-loop control to be realized, enterprises must be repeatedly reassessing the enterprise with some level of regularity. Yet, as

the case studies show, enterprises often fail to achieve this and cite the substantial time commitments of assessment as a primary factor for not reassessing the enterprise on a more regular basis (Abdimomunova, 2010). In one case study, a large aerospace enterprise reported, "the LESAT process took about six months. It was too time intensive, almost like pulling teeth, to get people to devote the time to complete the questionnaire. We have decided to stop using it partly because of that time commitment." They were an extreme case, but are nonetheless illustrative of the time investment required to coordinate the calendars of enterprise leaders and to complete assessment training, facilitation, scoring and interpretation.

There are two possible options for decreasing this time investment. On one hand, the number of participants could be decreased, but this goes against the intentions of the assessment as a self-assessment and the resulting benefits related to discussion, consensus building and measuring group cohesion (Nightingale & Mize, 2002). The other option is to try to decrease the time required to actually complete the assessment by reducing the number of practices. The full version of LESAT contains 54 practices, which compose a sixty-page questionnaire. To truly do justice to scoring and commenting on each practice is a substantial time investment that easily spans several hours, especially as each practice has a definition, descriptions of five levels of performance, and indicators. If the number of practices can be reduced without sacrificing the insights and knowledge gains that emerge from the assessment, then regular reassessment may become more appealing to enterprises.

With this motivation, we sought to find a methodology for reducing the number of practices scored in LESAT. This approach of reducing assessment inputs has appeared in other fields, especially medical surveys and assessments (Abdel-Khalek, 2006; Anderson, Gerbing, & Hunter, 1987; Sharp, Goodyer, & Croudace, 2006) where questions are reworded to better

capture the latent underlying construct. For enterprise assessment, this is more difficult as the latent construct is more abstractly defined (Denison & Mishra, 1995) – many have an idea of what enterprise excellence is, but few have a sufficiently objective vantage point to discern those qualities within their own enterprise. For this reason, rather than trying to define broader assessment criteria, we try to build a mathematical model to infer information about all practices based on responses to select subset. Respondents are then asked follow-up questions to confirm the predictions. In this way, an abridged assessment allows respondents to review those practices that are both most indicative and most relevant to their specific enterprise ensuring sufficient opportunity to gain crucial enterprise knowledge from the assessment process without reviewing all practices.

3.6. Summary

Throughout this chapter, we have reviewed the LESAT assessment process, several strategies and tools for interpreting results and the overarching motivations for creating an abridged assessment tool.

In order to maximize assessment benefits organizations need to watch out for common mistakes and threats to assessment validity. Assessment that accounts for organizational context and environmental factors is more likely to ensure that it fulfills specific needs of the organization. Consistent process, choice of respondents, scope of assessment and elimination of respondents' potential biases will address some threats to assessment validity.

A single LESAT assessment provides details about organizational strengths and weaknesses, as well as insights on existing communication structures and group cohesion. When LESAT becomes part of an ongoing, reassessment cycle, additional benefits are gained allowing for benchmarking, cross-enterprise comparison and analysis of temporal trends as a leading

indicator for transformation progress. All these tools combine to provide insights related to a lean transformation plan, in the form of measuring existing success and judging key areas of focus. Yet the secondary benefits, which are often overlooked, can be even more valuable. Use and analysis of LESAT can provide the insights and data necessary to drive needed enterprise behavior and motivate good decision-making. In addition, LESAT helps boost lean knowledge and understanding, all while aligning key personnel around the transformation plan, by providing them a clear understanding and basis for the direction of the transformation plan. The first-order and second-order benefits around LESAT make it a powerful tool for guiding the enterprise transformation process, especially when used in an integrated closed-loop control cycle where LESAT provides the primary driving force for improvement planning (Hallam, 2003).

These benefits from ongoing assessment make it ever more important to find a reasonable strategy for regular reassessment. The high number of practices in LESAT and the need to set aside time among a diverse group of enterprise leaders means many enterprises do not reassess, or if they do, without sufficient frequency to truly capitalize on the ongoing benefits. This unfilled need for reassessment motivates the creation of abridged assessment tools with fewer inputs, allowing the time requirements to be substantially reduced. In the next chapter, we discuss the methodology for actually creating an abridged assessment tool.

Chapter 4. Methodology

This chapter establishes a generalized methodology for creating an abridged assessment tool that can be used in lieu of the full assessment. An abridged assessment tool has the potential to alleviate key time and resource challenges associated by the assessment process, by reducing the number of inputs required. The abridged assessment is not meant to take the place of the full assessment, but simply create a complementary tool, which can be used for more frequent reassessment or can be used in resource-scarce situations.

Briefly, the methodology is a multistep process beginning with creating a predictive model to expose interconnection between assessment inputs. Yet, rather than solely depending on the outputs of the predictive model, the predictions are combined with a predetermined selection criteria to identify potentially high-priority follow-up questions. In this way, the predictive model acts as a tool for dynamically selecting a set of follow-up questions that are most relevant to the enterprise being assessed. The methodology is summarized in Figure 8. Subsequent sections will address each step in the methodology.

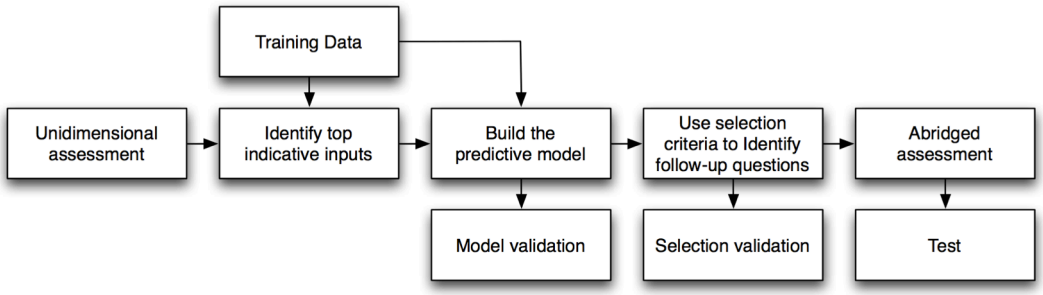


Figure 8 – Methodology for creating an abridged assessment

All steps in this methodology require having sufficient data from the full assessment. The data is initially used for assessment analysis, but also serves in the training of the predictive

model, midprocess validation of results and the testing of the overall abridged assessment. As a result, the importance of the data necessitates having a diverse collection of samples reflecting a range of potential assessment scenarios. Having more data enables the model to produce more accurate predictions and, as a result, include fewer inputs without sacrificing accuracy. In addition, to ensure objective testing of the abridged model, a portion of the data should be excluded from the training and validation process, in order to allow overarching testing of the full abridged assessment.

4.1. Unidimensional Assessment

This methodology is valid for unidimensional assessments, an assumption that is important for the construction of a predictive model. Unidimensionality refers to an assessment where all inputs are related by an underlying latent factor (Furr & Bacharach, 2008). Even though the common factor is intangible, immeasurable or abstract, its existence suggests that assessment inputs are linked and therefore can be mathematically inferred through a predictive model. Unidimensional surveys have long been prominent in medical practice, where diverse questions about wellbeing or mood can be used to measure latent psychological qualities (Sharp et al., 2006). This quality is transferrable to enterprise assessment, where a diverse range of questions about independent activities or behaviors are all linked by a common factor – broadly conceptualized by terms such as “enterprise excellence”.

Unidimensionality provides a valuable starting place for the methodology as it ensures that all inputs to the assessment are related to a common theoretical construct of interest. As a result, a subset of the inputs should still coarsely measure that same construct of interest. Since the construction of the predictive model (described in the subsequent section) uses least squares regressions, the calculation of model parameters can still fit inputs from a multidimensional

assessment. The risk of applying this methodology to a data from a multidimensional assessment is that secondary or tertiary latent factors may not be adequately represented by the subset of inputs used by the predictive model.

As a result, multidimensional assessments should be carefully analyzed to determine the latent factors of interests. In some cases, additional dimensions may simply reflect poor assessment design, as the additional dimensions reflect irrelevant or undesired information (Venkatraman, 1989). In such a scenario, the tool must be reevaluated to disentangle the underlying theoretical construct of interest from the additional factors being measured. Alternatively, other multidimensional assessments may reflect an intentional need to capture diverse and independent factors of performance. In this case, the assessment inputs should be grouped according to the factors being measured and the methodology should be applied for each unidimensional group. By doing so, the abridged assessment model will likely reflect all desired dimensions.

Verifying unidimensionality can be done in one of two ways: either using principal component analysis (PCA) or Cronbach's alpha. Principal component analysis is a transformation that projects possibly correlated variables into uncorrelated variables called the principal components. The first principle component has as high a variance as possible, and subsequent components then are calculated to achieve the highest variance possible while remaining orthogonal (uncorrelated) to the previous components. In this way, PCA reduces the dimensionality of data by turning a large number of interrelated variables into uncorrelated, order variable (Jolliffe, 2002). The order factors calculated by PCA show latent factors and their relative impact, allowing confirmation of unidimensionality of all inputs. By finding the eigenvalues of the correlation matrix for the input data, the relative impact of each principal

component can be ascertained (Furr & Bacharach, 2008). If the first eigenvalue is substantially larger than subsequent eigenvalues, it suggests a single latent factor accounts for much of the variation in inputs and hence that the assessment is unidimensional.

A secondary method for confirming unidimensionality is to calculated Cronbach's alpha, a psychometric tool used to measure internal consistency and reliability (Cronbach, 1951). The coefficient is calculated as a ratio of two variances: the variance of each component input, and the variance across respondents. The resulting ratio is then normalized based on the number of components, and subtracted from one, providing a final coefficient between zero and one. As Cronbach's alpha approaches one, it indicates internal consistency – that is, an increased intercorrelation among component inputs (Anderson et al., 1987). An alpha above 0.85 is indicative of first-factor saturation, meaning that a predominant latent factor accounts for variation in scores (Cortina, 1993). A sufficiently high Cronbach's alpha suggests unidimensionality, but it is insufficient evidence on its own and hence is used in addition to the principal component analysis (Gardner, 1995).

Once the principal component analysis and Cronbach's alpha have been calculated, it is possible to ascertain whether the assessment in question is unidimensional, as indicated by a relatively dominant first principal component and a Cronbach's alpha above 0.85. Once the assessment has been determined to be unidimensional, or once the assessment has been subdivided in to unidimensional parts, it is possible to begin building the predictive model that will allow extrapolating information from a subset of inputs regarding the underlying latent factor.

4.2. Step 1: Identify Top Predictive Practices

The first step to creating the prediction model requires an iterative process of feature selection to identify a subset of assessment inputs that best enable calculating predictions. In order to identify the subset of inputs, a nested subsets feature selector implementing a greedy forward selection is used in combination with a linear least squares learning model (Guyon & Elisseeff, 2003). By calculating iterative linear least squares regressions, the ideal subset of inputs can be efficiently identified while substantially reducing prediction model errors. The iterative process goes through four key steps:

1. A candidate input is selected from the assessment, which does not yet appear in the subset of inputs used by the predictive model.
2. Using the existing subset of inputs as well as the candidate input, the linear least squares prediction model is calculated for all other inputs minimizing the mean square error (MSE) for other inputs.
3. Using k-fold cross-validation, the new predictive model is scored in terms of the mean square error.
4. Steps 1-3 are repeated for all possible candidate inputs. The lowest k-fold cross-validation mean square error is then compared with the existing predictive model, from previous iterations. If the addition of another feature to the model does not result in an improvement to the mean square error, then the iterations are terminated (providing a stopping condition).
5. The candidate input that produced the lowest k-fold cross-validation mean square error is then added to the subset of inputs.

Through this process, a subset of inputs can be selected that efficiently and accurate predict the other inputs. By using k-fold cross validation, the iterative process is terminated prior to overfitting (Moore, 2006).

Although this methodology includes prescriptive guidance regarding algorithms and tools for feature selection, this methodology is algorithm-agnostic. Alternative feature selection algorithms can be implemented based on the specific needs of the assessment. For example, high dimension data might benefit from a feature-selection algorithm that is further optimized. Simply using Gramm-Schmidt projections, the process of feature selection can be greatly expedited (Stoppiglia, Dreyfus, Dubois, & Oussar, 2003). Given the small number of inputs in most assessment tools and the computational tractability of the brute-force, iterative process describe above, the methodology presented in this paper does not reflect the many alternative implementations that may include performance benefits (Somol, Novovičová, & Pudil, 2010).

4.3. Step 2: Build the Predictive Model

Using the subset of inputs that are able to minimize squared residuals when predicting other inputs, the next step is to build a predictive model that produces estimates for all other inputs based on those in the known subset. The predictive model is generated using the same learning algorithm as in step two of the last section – least squares regression.

As with the feature selection process, this methodology assumes a simple learning algorithm that is ideally suited to most assessments, but is not dependent on this specific choice. Alternative learning algorithms can be explored to find an appropriate tradeoff between accuracy, overfitting and processing time. For example, we explored neural networks as an alternative to linear least squares, but found the tradeoffs in training times did not enable suitable increases in prediction accuracy to warrant further exploration.

The predictive model should also integrate structural or contextual information regarding the assessment to improve accuracy. For example, in assessments that require respondents to make tradeoffs or to rate inputs relatively, the degrees of freedom for each estimated input decrease. By integrating these constraints into the predictive model, the accuracy can be increased. The opportunity to make structural or contextual changes to the predictive model depends on the assessment.

Model Validation

Now that a predictive model has been created, an initial validation step can occur, in order to ascertain the accuracy of estimates produced by the predictive model. Using data from full assessment, estimates can be calculated based on the indicative inputs and then compared with the known values. Mean square error provides a useful metric for understanding the accuracy achieved by the model, as well as for making comparisons between different modeling algorithms.

4.4. Step 3: Use Selection Criteria to Identify Follow-up Inputs

The abridged assessment is not meant to solely depend on the predictions as a source of information. Instead, the follow-up inputs are selected to confirm the estimates from the prediction model. After a respondent provides the subset of predictive inputs needed, the prediction model is used to calculate estimates for the remaining inputs. Then by applying predetermined selection criteria to the estimates produced by the prediction model, follow-up questions are selected in order to address potential inputs of concern. The decision regarding selection criteria is flexible and should align with the desired assessment outputs. A specific selection criteria need not be bundled with the abridged assessment, but rather, can be changed

based on the needs of the enterprise or organization implementing the assessment process. The selection criteria for follow-up questions can be either generic (highest or lowest performing inputs) or can be contextually tailored to the enterprise completing the assessment (largest change since last assessment).

Selection Validation

An important validation step for the abridged assessment is measuring its ability to successfully identify high-priority follow-up inputs. High-priority follow-up inputs are defined as those falling into a predetermined quantile based on the selection criteria. For assessments with a small number of practices, this may be defined as the top quartile, while assessments with many practices may define a stricter cutoff (e.g., top tenth percentile). The goal is to ensure that sufficient follow-up inputs are assessed to gain the desired, actionable information from the assessment process. The validation step provides an opportunity to calculate the number of follow-up questions that should be collected to ensure that the high-priority practices are assessed. Because of the inherent coarseness of predictions from the model, high-priority practices may not always be identified and an acceptable cutoff rate must be defined. For this methodology, we use a 95% cutoff rate. That is, enough follow-up inputs should be collected to ensure that the respondent assesses 95% of high-priority practices.

To validate the process and calculate the necessary number of follow-up inputs, full assessment data is used, where high-priority practices can already be identified. The actual high-priority practices are compared with those identified by the predictive model. Using the full data set, a threshold number of follow-up inputs can be identified to ensure that, on average, a respondent assesses 90% of high-priority practices, minimizing dependence on predicted scores for key assessment outputs.

One complicating factor for selection validation involves low granularity assessments, where separating the highest priority inputs is confounded by a lack of data granularity in responses. If practices cannot be uniquely ordered, it may not be possible to distinguish a specific subset as the high-priority practices based on selection criteria. For such scenarios, two possible remedies exist. The first option is that the abridged assessment should be judged based on its inclusion of a sufficient number of any of the high-priority practices. This calculation is achieved by figuring out which non-unique subset of high-priority practices is most represented in the follow-up questions, and using that non-unique subset to calculate the percentage as normal. The second option is to reverse the validation. Instead of calculating the number of high-priority practices identified, the validation can be based on the percentage of follow-up questions that meet the selection criteria. These two options provide two alternatives for overcoming low-granularity assessment data in the validation phase – the second option is a higher bar, but allows the abridged assessment to be more inclusive of potential high-priority practices.

4.5. Step 4: Abridged Assessment

Having built a predictive model and identified both potential selection criteria and the corresponding number of follow-up questions necessary, the abridged assessment can be used. The process of using the assessment requires two phases, one where respondents provide the inputs necessary for the predictive model and one where the respondent provides inputs based on selected follow-up questions. The number of inputs in each phase depend on both the predictive model and the desired number of follow-up questions respectively. The assessment process is shown in Figure 9.

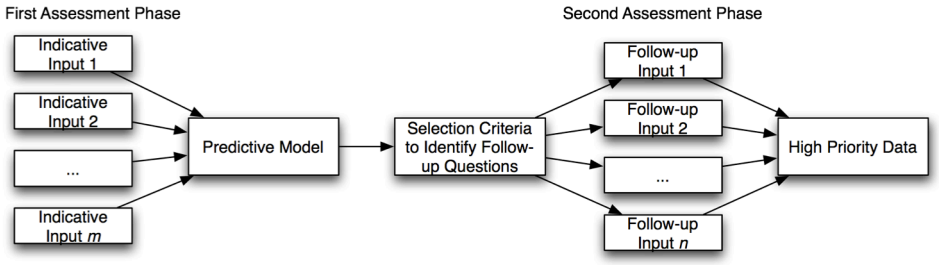


Figure 9 – Process for performing abridged assessment

Because of the two-phase assessment process, the abridged assessment is best suited to a digital media (online or software based assessment), but can also be done in two rounds (where responses to the first phase are used to generate the second phase). In addition, because of the design of the predictive model, all inputs in the first phase must be provided.

Test the Abridged Assessment

To test the process, full assessment data are used, where high-priority practices can already be identified and compared with what the abridged assessment would have identified as high-priority practices. As with any statistical learning algorithm, the data used in the testing step must be distinct from the data used in the previous training steps to ensure that the learning is not specific to the noise in the training data. In order to measure the effectiveness of the abridged assessment, the number of high-priority follow-up inputs identified is compared to a random selection process. That is, does the abridged assessment more accurately select high-priority inputs then a process of randomly selecting follow-up inputs. Equation 1 shows the probability density function for the random process, where n is the total number of practices, g is the number of follow-up questions asked, d is the desired number of high-priority inputs that should appear in the follow-up questions and a is the actual number of high-priority inputs that appear in the follow-up questions.

Equation 1 - Probability density function

$$f(a;d,g,n) = \frac{\binom{n-d}{g-a}\binom{d}{a}}{\binom{n}{g}}$$

for integer values subject to $0 \le a \le d \le g \le n$

The probability density function provides a reference point by which the abridged assessment can be compared. The goal is to ensure that the abridged assessment is able to select follow-up questions that are significantly more likely to be high-priority practices then if it were a random selection process. Once follow-up questions have been identified and compared to the high-priority inputs for the testing data, the probability density function should be used to calculate a *p* value for each data point in the test data; *p* represents the probability that the selection of high-priority practices is random. A well-designed abridged assessment will have a *p* value consistently below 0.05. This evaluation of the test data can be expressed graphically by plotting a histogram of the effectiveness of the abridged assessment against the probability density function described.

It is worth noting that certain high-priority inputs may be included in the set of indicative practices of the first phase of the abridged assessment. For the purposes of this testing, high-priority practices should only include those practices that do not appear in the indicative set from the first phase.

Should the *p* value calculated during the testing phase be too high, the abridged assessment design process can be repeated in ways designed to increase the effectiveness of the predictive model. Potential changes include: more training data, better learning algorithms for

the predictive model and/or increasing the number of follow-up questions asked. By implementing such improvements, the p value should decrease demonstrating the increased efficacy of the abridged assessment.

Next, we present an application of this methodology on LESAT.

Chapter 5. Applying the Methodology: LESAT

Having established a methodology for creating, validating and testing an abridged assessment, we can now apply this methodology to an enterprise assessment tool. The Lean Enterprise Self-Assessment Tool (LESAT) introduced in Chapter 2 and described in more detail in Chapter 3 is an ideal candidate for the abridged methodology. The assessment has fifty-four practices that each must be scored twice, once based on the current performance of the enterprise and once based on the desired performance after a predetermined transformation timeline (the full list of practices appears in Appendix A: LESAT Structure). As a self-assessment meant for enterprise leaders, it is time intensive requiring the input and coordination of top individuals within the enterprise (Nightingale & Mize, 2002), which has limited its adoption (Abdimomunova, 2010). By creating an abridged LESAT, the goal would be to significantly reduce the number of questions each respondent must answer.

5.1. Data

Training, validating and testing an abridged LESAT requires a sizable collection of data points from completed assessments. For training the abridged LESAT, we analyzed the LESAT data from 271 respondents from 24 companies (each company having between 5 and 19 respondents). For each respondent/practice pair, the three score states were treated as independent – the current performance, the desired performance and the gap. Although the gap can be derived from the first two inputs (subtract the current performance from the desired performance), creating separate predictive equations for gaps decreases the degrees of freedom between other inputs and helps improve accuracy of the predictive model (as discussed in section

4.3). The data were collected by PhD student Cory Hallam for use in his dissertation from 24 enterprises in the US and UK aerospace industry (Hallam, 2003). Senior leadership at the selected organizations completed the assessment under Cory's guidance, and with proper training in the assessment tool. Some respondents (n=143) omitted scores for one or more practices, either as an error or because they did not feel sufficiently informed to assess the specific practice. Throughout the subsequent steps, the data are appropriately down selected where needed inputs are not present. In order to maximize the benefit from the training process, this is done conservatively – only dropping data points when the omitted inputs are required for the learning algorithm or for the selection of follow-up practices.

At the beginning of the process, 15% of the dataset was randomly selected and set aside for testing purposes.

5.2. Unidimensionality

Before applying the abridged assessment methodology, it is important to verify that LESAT is a unidimensional assessment with sufficiently high inter-factor correlations – that is, that the assessment measures one primary latent factor. When LESAT was designed, each practice was intended to measure an independent practice illustrative of a lean enterprise (Murman et al., 2002; Nightingale & Mize, 2002). Despite the fact that practices were selected with intentions of creating independent indicators, it seems likely that they are tied to a common latent factor related to a measure of enterprise "leanness."

This is confirmed by principal component analysis, which was performed for each input state (current, desired and gap). Each state shows a dominant latent factor that accounts for the majority of the variance between respondents. This can be seen in Figure 10, which shows a

graph of the eigenvalues of the correlation matrix. The large value for the first latent factor in the graph is indicative of a unidimensional assessment.

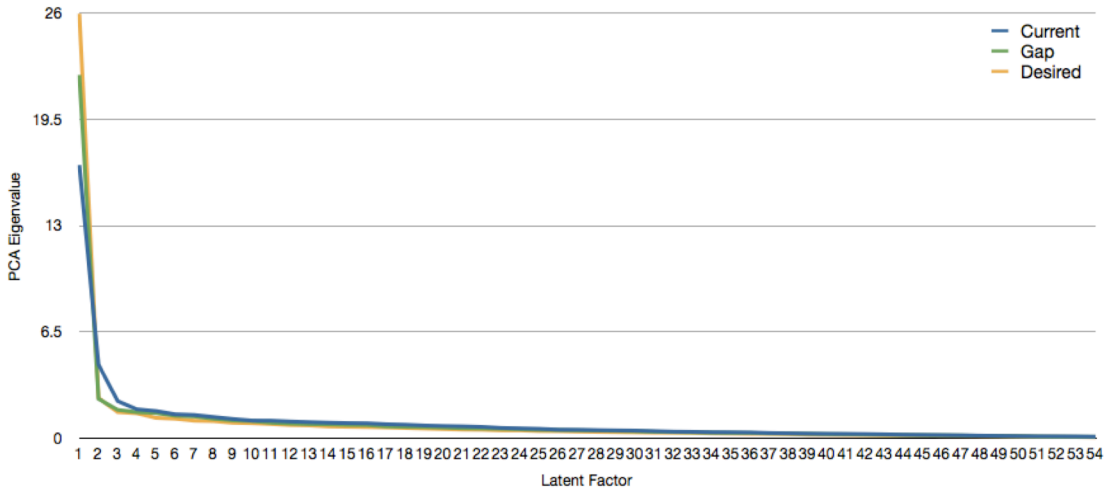


Figure 10 – Principal Component Analysis (PCA) for LESAT practices

Calculating Cronbach's alpha provides additional confirmation of the unidimensionality of the assessment. Alphas above 0.85 indicate high internal consistency and first factor saturation, and as Table 6 shows, variation in LESAT inputs is predominantly accounted for by a single latent factor. Both the principal component analysis and the Cronbach's alpha indicate that LESAT is a unidimensional assessment that predominantly measures a single latent factor – presumably, enterprise "leanness." This step verifies that consolidating LESAT down into an abridged assessment will not risk jeopardizing secondary dimensions captured by the full assessment process.

Table 6 – Cronbach's alpha coefficient for LESAT

State	Cronbach's α
Current	0.956
Desired	0.979
Gap	0.984

5.3. Predictive Model Accuracy

To build the predictive model for the abridged LESAT, indicative practices were selected independently for each input state (current, desired and gap). Linear least squares estimators with 20-fold cross-validation were used in training the learning algorithm for the iterative selection process. As a result, three lists of most-indicative practices were identified and are shown in Table 7, in the order that they were added to the predictive model.

Table 7 – Top indicative practices for each input state

Predicting Current Scores	Predicting Desired Scores	Predicting Gap Scores
1. I.G.2 Monitoring Lean Progress	1. I.F.1 Development of Detailed Plans Based on Enterprise Plan	1. II.E.1 Utilize Production Knowledge and Capabilities for Competitive Advantage
2. I.D.1 Enterprise Organizational Orientation	2. II.F.3 Enhance Value of Delivered Products and Services to Customers and the Enterprise	2. I.C.3 Designing the Future Value Stream
3. II.A.3 Provide Capability to Manage Risk, Cost, Schedule and Performance	3. II.D.2 Optimize Network-Wide Performance	3. II.F.4 Provide Post Delivery Service, Support and Sustainability
4. I.B.3 Lean Enterprise Vision	4. I.C.4 Performance Measures	4. I.G.3 Nurturing the Process
5. II.D.1 Define and Develop Supplier Network	5. III.B.2 Common Tools and Systems	5. II.F.2 Distribute Product in Lean Fashion
6. III.A.4 Enable the Lean Enterprise with Information Systems and Tools	6. II.F.2 Distribute Product in Lean Fashion	6. I.B.2 Senior Management Commitment
7. I.D.4 Employee Empowerment	7. I.C.3 Designing the Future Value Stream	7. II.A.4 Allocate Resources for Program Development Efforts
8. I.C.4 Performance Measures	8. II.C.1 Incorporate Customer Value into Design of Products and Processes	8. II.F.1 Align Sales and Marketing to Production
9. II.F.3 Enhance Value of Delivered Products and Services to Customers and the Enterprise	9. I.B.2 Senior Management Commitment	9. I.D.4 Employee Empowerment

Although the practices are listed in the order that they added to the predictive model, it is important to caution against interpreting the order of practices. The first practice may be the most

indicative, but subsequent practices are selected not for their indicative qualities but instead for their ability to account in errors in the existing predictive model. It is worth pointing out that there is overlap between the three lists. For example, "Enhance Value of Delivered Products and Services to Customers and the Enterprise" and "Employee Empowerment" appear in two of the lists of indicative practices.

The lists of indicative practices only include the top nine most indicative practices, although additional practices were identified before the cross validation triggered the stopping condition (i.e., prevented over fitting). It is worth noting that adding additional practices to the predictive model only generates marginal decreases in the mean square error, as illustrated by Figure 11.

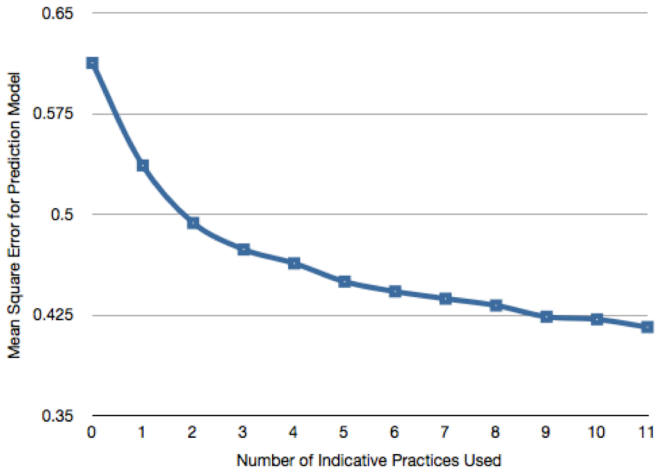


Figure 11 – Mean square error as number of indicative practices increases

Based on the list of identified practices and the marginal decreases in mean squared error, we can build a predictive model for the whole assessment by selecting the top two practices from each category, and combining them into a single predictive model (the first two indicative practices achieve the majority of the decrease in mean square error). The combined predictive model uses the following practices:

1. I.C.3 Designing the Future Value Stream
2. I.D.1 Enterprise Organizational Orientation
3. I.F.1 Development of Detailed Plans Based on Enterprise Plan
4. I.G.2 Monitoring Lean Progress
5. II.E.1 Utilize Production Knowledge and Capabilities for Competitive Advantage
6. II.F.3 Enhance Value of Delivered Products and Services to Customers and the Enterprise

It is worth noting that all practices are from Section I and II of LESAT and the number of indicative practices from each section is roughly proportional to the total number of practices in that section, but no practices from Section III appear in the predictive model. Although the predictive model would function well with questions from Section III as well, it is clear that they were not necessary for the abridged LESAT. This is in line with existing research on the interconnection between LESAT sections, which suggests that improvements in Section I and III are correlated with improvements in Section II (Hallam, 2003). As a result, it is not unreasonable that the scores from Section I and II provide sufficient information to infer Section III scores.

A combination of the inputs listed above was used to train the combined predictive model. The prediction errors can be further decreased by utilizing assessment-specific knowledge, such as the range in possible values and the relationship between the current/desired state and the gap state. The range of values constraints possible predictions to between one and five, while the relationship between the three states (current, gap and desired) helps decrease the degrees of freedom in the predictive model. By imposing the constraint that the current score plus the gap should always equal the desired score, the errors in the predictions are further decreased. The newly trained and constrained prediction model achieves estimate errors shown

in Table 8 (MSE for each practice appears in the Appendix C: MSE for Final Predictive Model). Since each practice has its own predictive formula, the model is better at predicting some practices. In order to provide a range of the prediction errors for different practices, the table shows the most and least accurately predicted practices in addition to the average MSE for all practices.

Table 8 – Mean square error for complete prediction model

	Predictive Model		
	Average MSE for all practices	**Minimum MSE (most accurate)**	**Maximum MSE (least accurate)**
Current	0.1938	0.122 (Practice #29)	0.357 (Practice #31)
Desired	0.189	0.128 (Practice #29)	0.287 (Practice #44)
Gap	0.222	0.159 (Practice #9)	0.361 (Practice #5)

Now that a prediction model has been generated and shown to have low error rates for predicting other LESAT practices, it is informative to apply the prediction model to the challenge of identifying potentially high-priority practices.

5.4. Follow-up Practice Identification

The next step in building an abridged LESAT is to identify the number of follow-up questions to ask, and measure the effectiveness of using the predictive model to select follow-up questions. Inherent in this step of the methodology are decisions based on value judgments pertinent to the assessor based on the desired outputs from the process. We have sought to create a generic and generally viable version of the abridged LESAT for use in reassessment, and have made our decisions accordingly, but these are not inherent "the right" choices. First, we have sought to identify six high-priority practices in the follow-up practices. Given the need for enterprise focus during transformation, a small number of high-priority practices should be

sufficient, and six was chosen to mirror the number of indicative practices (asked in phase one of the assessment process). Second, we are focusing on low current scores as selection criteria. A viable alternative abridged LESAT would select follow-up practices based on largest gaps, but that is less important in reassessment (as reassessment focuses heavily on measuring progress or key enterprise limitations).

Based on the iterative approach described in the methodology (see section 4.4), asking twelve follow-up questions is sufficient to identify 90% or more of the high-priority practices. Figure 12 is a histogram illustrating the effectiveness of the predictive model at selecting follow-up questions. The green line illustrates the probabilistic distribution if follow-up questions were selected by a random process (based on Equation 1 from page 101), while the blue bar chart is a histogram represents the actual number of data points where the corresponding percentage was achieved (e.g., for 109 of the LESAT respondents, the abridged LESAT would have assessed all six high-priority practices with the follow-up questions).

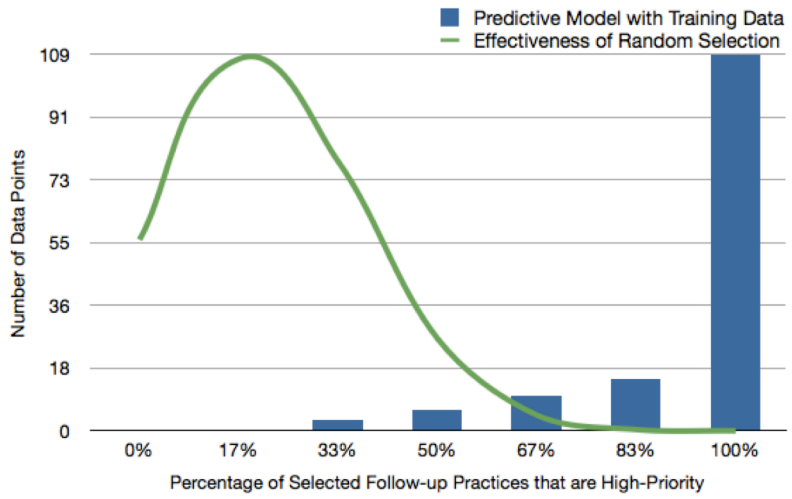


Figure 12 – Histogram showing percentage of follow-up practices that are high-priority practices for the training dataset

Interpreting Follow-up Questions

Although the tool selects distinct follow-up questions for each respondent based on their responses to the initial six practices, there are interesting patterns in which follow-up questions are selected. For example, there are nine practices, which never appeared in the follow-up questions selected for any of the respondents in the training dataset. This suggests that these practices tend not meet the selection criteria used (i.e., these nine practices tend to not have low current performance scores). At the same time, some practices appear in the follow-up set regularly, suggesting that they represent more difficult areas for enterprises to achieve improvement. Alternatively, these patterns may suggest inconsistency in the scales for the outlying practices (that is, those practices that appear more often may be defined in such a way that moving from a level two to a level three is easier than for other practices). A list of both the most commonly and least commonly appear follow-up questions appears in Table 9.

Table 9 – Most and least common follow-up practices based on applying the predictive model to the training dataset

Least Common Follow-up Practices	Most Common Follow-up Practices
• I.A.1 Integration of Lean in Strategic Planning Process • I.A.3 Leveraging the Extended Enterprise • I.D.3 Open and Timely Communications • II.A.3 Provide Capability to Manage Risk, Cost, Schedule and Performance • II.A.4 Allocate Resources for Program Development Efforts • III.A.4 Enable the Lean Enterprise with Information Systems and Tools • III.A.5 Integration of Environmental Protection, Health and Safety into the Business • III.B.1 Process Standardization • III.B.2 Common Tools and Systems	• I.B.3 Lean Enterprise Vision • I.B.4 A Sense of Urgency • I.D.7 Lean Change Agents • I.E.1 Enterprise-Level Lean Transformation Plan • II.A.1 Leverage Lean Capability for Business Growth • II.B.2 Utilize Data from the Extended Enterprise to Optimize Future Requirement Definitions • II.D.3 Foster Innovation and Knowledge-Sharing Throughout the Supplier Network

5.5. Testing the Abridged LESAT

Thus far, training and validation have used a shared dataset. In order to ensure the effectiveness of the abridged LESAT, it must be tested with novel data (data other than that used for training). Using the testing dataset, the ability of the predictive to select relevant follow-up questions can be assessed in the same way as for training data. Figure 13 is a histogram illustrating the effectiveness of the predictive model at including high-priority practices in the selected follow-up practices. As with the previous figure, the green line represents the probabilistic distribution if a random selection of follow-up practices occurred, while the blue columns represent the number of testing data points where the corresponding percentage of high-priority practices were included in the follow-up practices (e.g., for 11 of the LESAT respondents, the abridged LESAT would have assessed all six high-priority practices during the follow-up questions).

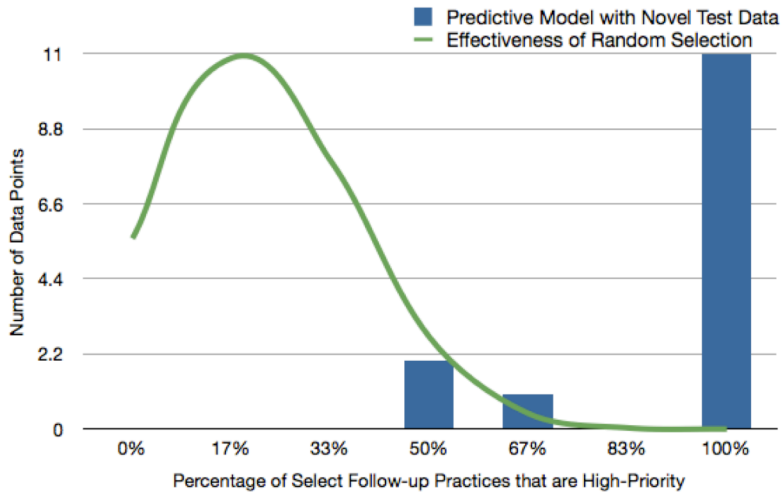


Figure 13 – Histogram showing percentage of follow-up practices that are high-priority practices for the testing dataset

By examining the graph, it appears that the abridged LESAT is effective at assessing high-priority practices. By comparing the distribution of the random selection process with the

predictive model, it is possible to demonstrate that the abridged LESAT has a statistically significant ability to select follow-up questions that correspond high-priority practices (p = 0.0133).

5.6. Conclusions

Using the methodology described in the previous chapter, we have created, validated and tested an abridged version of the Lean Enterprise Self-Assessment Tool. Instead of asking all 54 practices that appear in the full assessment, the abridged assessment allows valuable assessment results to be gained using only 18 practices. Six practices compose the inputs to the predictive model, which then drives the selection of twelve follow-up questions. The abridged LESAT includes 7.7 high-priority practices on average,[1] which is statistically significant when compared to a random selection of follow-up questions. This suggests that the abridged LESAT is a viable and accurate, but lower granularity version of the full assessment tool and has functional potential as a complement to the full assessment.

[1] This is higher than the desired six high-priority practices, because of the low-granularity in LESAT results, making the ranking of practices non-unique.

Chapter 6. Applying the Abridged LESAT

In the previous chapter, we created and validated an abridged version of LESAT that allows an enterprise to capture high-priority practices based on a predetermined selection criterion (in the example case, based on the lowest current scores) without the full burden normally associated with completing a LESAT assessment. This allows organizations to overcome one of the most substantial deterrents from adoption of a cyclical assessment process and from regular reassessment. By decreasing the number of items in the assessment by 66%, respondents will better able to find time to perform assessment and leadership will have fewer reservations about prioritizing assessment as a value-added task.

Despite both the benefits of the abridged assessment and the similarities between the abridged LESAT and the full LESAT assessment, the abridged assessment cannot simply be dropped into the original assessment process. The unique nature of the abridged LESAT suggests a distinct role in the assessment process, and requires changes to both the process and to the interpretation of results.

6.1. Implications and Role of an Abridged Assessment

It is important to acknowledge both the implications and role of the abridged LESAT. The abridged assessment has strengths and limitations that are distinct from the full assessment, and these must be considered carefully before opting to use the abridged assessment.

Implications

The creation of an abridged version of LESAT has implications for the broader assessment tool. Primarily, it confirms the unidimensionality of the assessment confirming that

LESAT does indeed measure a distinct latent factor. Conceptually, this is a valuable insight as it confirms that the seemingly independent practices are all related by underlying connections that can be labeled either "leanness" or more generally "excellence." A similar results has been confirmed by explicit research looking at the links between sections, which confirmed that improvements in the mean of section one practices led to improvements in both section two and three (Hallam, 2003). But unidimensionality goes a step further, showing that all practices are interrelated and form a single dimension that can be thought of as performance.

Another interesting and important implication of the abridged LESAT is that it is dominated by practices from Section I (leadership and lean transformation) and Section II (lifecycle processes), while Section III (supporting processes) is almost always absent (see Table 10). All of the most predictive practices are drawn from Section I and II, and the common follow-up practices (see Table 9 in the previous chapter) also come from Section I and II. In fact, four of the practices from Section III were never included in the follow-up practices for the datasets used for training and validation of this model. There are several potential explanations that probably all contribute to this absence. First, the dominance of section one in the abridged LESAT is reflective of research that emphasizes the importance of leadership in both enterprise performance and in transformation (George, 2006; Kotter, 1995). Second, practices from section three tend to act as an important stepping-stone that connects leadership improvements to lifecycle improvements (Hallam, 2003). That is, supporting processes are dependent on leadership and are reflected by the performance of the lifecycle processes. As a result, it is easier for the predictive model to make estimations for practices in Section III. Third, some of the practices in Section III have less variability across organizations. For example, the practice III.A.5 "Integration of Environmental Protection, Health and Safety into the Business" has

consistently high ratings across respondents. The emphasis on safety and caution in a manufacturing environment reduces the variability in responses, meaning it often does not add to the predictive model nor is it a high-priority practice.

Table 10 – Section representation in LESAT and the abridged LESAT

		Full LESAT	Predictive Model	Common Follow-up Practices	Abridged LESAT
Section	**1. Leadership and Enterprise Transformation**	28 (52%)	4	8	12 (67%)
	2. Lifecycle Processes	18 (33%)	2	3	5 (28%)
	3. Supporting Processes	8 (15%)	0	1	1 (6%)
Total		54 (100%)	6	12	18 (100%)

Percentages may total more than 100% due to rounding. The abridged LESAT represents the sum of the predictive model and the most common follow-up practices.

One final implication gained from applying the abridged assessment methodology to LESAT is that five-level scale makes it difficult to differentiate practices. Although it is not immediately clear in the results chapter, the lack of granularity in individual responses makes it difficult to select a unique set of high-priority practices. When looking through the training dataset to try and identify six high-priority practices, on average over eighteen potential high-priority practices were found. This lack of granularity is a consequence of having a five level scale for scoring practices, and is further confounded by the fact that level five is representative of a world-class organization, and hence, is rarely used (limiting practices to a potential scale of four). This lack of granularity is not necessarily a bad thing (any increase in granularity may simply be an artifact of a more complex scale of performance), but it does make it challenging to differentiate high-priority practices at the individual level.

Threats to Validity and Limitations of the Abridged LESAT

The newly created abridged LESAT has some key limitations that are important to understanding the role and strengths of the abridged LESAT as an assessment tool. Although the predictive model is effective at selecting follow-up questions based on predetermined selection criteria, the abridged LESAT is limited by the fact that follow-up questions are only based on a specific set of criteria. This requires that the user of the abridged tool know what sort of information is desired from the assessment process. Because follow-up question selection is based on a single set of follow-up criteria, the assessment will inherently have less broad applicability. For example, if the abridged LESAT uses "lowest current scores" as the selection criterion, it will be hard to make decisions or gain insights regarding the largest gaps. For this reason, the abridged LESAT is most fruitful in relation to the selection criteria used.

From this, we can infer one of the larger limitations of the abridged LESAT: it requires knowledge and familiarity with the full assessment tool. For an enterprise to really benefit from the abridged LESAT, it should be used as a simply a reassessment tool. The initial assessment should utilize the full assessment, allowing respondents to be exposed to all practices and to become familiar with the evaluation and interpretation process. From this first exposure to LESAT, the facilitator can learn what selection criteria would best enable the abridged LESAT to provide usable outputs for the strategic planning process. In addition, the first exposure will allow assessment respondents to get a broad picture of the principles and practices included in LESAT. If respondents only took the abridged LESAT, they may lack the confidence that the full set of practices is inclusive undermining trust in the results of the assessment.

Another limitation of the abridged LESAT is that each respondent may assess a slightly different subset of practices. Certain practices will have no score (besides the predicted value), and other practices may have only one or two scores. Variation in the selection of follow-up

practices can provide added insights (see 6.3, a discussion of methods for interpreting the abridged assessment results), but it also means that some interpretation or comparison strategies may not be available. For example, two distinct subgroups of respondents could not be compared if only one or two respondents scored a specific practice. In addition, this undermines the insights gained from the variance of scores for practices with few respondents. Although variance can still be used to measure group cohesion for the six practices in the predictive model and for practices scored by many of the respondents, it ceases to be valuable for those practices with only a few respondents.

A final limitation to the abridged LESAT is that respondents cannot skip any of the six practices presented in the first assessment phase, because these provide important guidance for the prediction model. Skipping practices would prevent the prediction model from selecting the most relevant follow-up questions. Of course, the six practices that appear in the first phase of the abridged LESAT are broadly applicable and should be relevant to all enterprises using the tool.

Although these limitations may impair specific implementations, there are more broad threats to the validity of the abridged assessment methodology. First and foremost, the abridged version of LESAT is dependent on the data that are used to construct the predictive model. A substantial dataset was used to train, test and validate the predictive model, but the dataset is based on companies in a common industry (aerospace) and with common interests (all companies had an association with the Lean Advancement Initiative, and hence an interest in lean business practices). In addition, the dataset can be viewed as a convenience sample (Hallam, 2003). Hence, the resulting predictive model and, therefore, the abridged assessment may not be fully generalizable. We have tried to account for generalizability in our methodology by using

test data from a different source (collected by different individuals, and from a defense industry enterprise), but more data is needed before the abridged assessment can be more broadly generalized.

Likewise, the abridged assessment is limited by the training methodology behind the predictive model. Selecting a method for predicting scores and training it is somewhat subjective. Although the mean square error metrics provide indications of the accuracy of the predictive model, it is impossible to confirm that underlying predictive model truly represents real-world associations between practices. For example, in the case of abridged LESAT, linear least-squares approximation was used to associate practices, but this may not fully capture relationships between the LESAT practices (a nonlinear model may have more accurately provided insights). Use of performance metrics (such as MSE and cross validation) help identify a strong and suitable underlying predictive model, but the decision is subjective and may not fully reflect the underlying relationship between assessment practices. These subjective decisions regarding the training methodology and model, as well as the resulting errors in the prediction model, can further compromise the resulting abridged assessment tool.

Strengths of the Abridged LESAT

The abridged LESAT is a valuable tool for enterprises that are familiar with the assessment process and are interested in expediting reassessment. Although the limitations presented in the last section make the abridged LESAT an insufficient replacement from the full assessment, it is a suitable substitute when it comes to reassessment. The abridged LESAT has only one third of the normal practices, and still captures high-priority practices based on a flexible set of selection criteria (the example used in the previous chapter was lowest current scores).

As a result, the abridged LESAT still offers six first-phases practices that are assessed by all respondents, and a mix of additional second-phase practices that represent an area of focus. The commonality in the practices assessed creates opportunities for in-depth analysis of results, and the selection of relevant follow-up practices creates the needed inputs to differentiate the core areas for enterprise improvement. The abridged LESAT can be applied in a way to track trends in the first-phase practices (allowing the abridged LESAT to act as a measuring stick of existing enterprise transformation), while the second-phase practices act as inputs into the formation of a coherent transformation plan to guide future transformation efforts. These potential uses make the abridged assessment ideally suited for the assessment process in cases where sufficient baseline data and knowledge have already been established.

6.2. Recommended Abridged Assessment Process

In section 3.1, the recommended LESAT assessment process was described. The five phase process for performing assessments to drive transformations, as developed by Abdimomunova (2010), remains the relevant for the abridged version of LESAT. Nonetheless, some changes are needed in order to ensure that the shortened assessment process does not detrimentally affect the results and outputs of the process. The following sections introduce some key changes that adapt the recommended LESAT assessment process to the abridged version of the assessment tool. The recommended changes are limited to the second (plan assessment), third (perform assessment) and fourth (evaluate assessment results) phase of the recommended process.

6.2.1. Changes to Phase Two: Plan Assessment

The second phase requires minimal adjustments to ensure that all respondents understand both the full assessment and what differs in the abridged assessment. In addition, certain decisions about decision criteria need to be made earlier in the assessment process, allowing the abridged assessment to be adapted to the needs of the enterprise.

Introduce the Tool to New Respondents. Since the abridged LESAT is designed for reassessment, it is assumed that most respondents will have previous exposure to the full assessment tool. For any respondents who are new to LESAT should be introduced to the full version, so that they are aware of the multitude of factors related to lean enterprise performance that are incorporated into the assessment process. Ideally, if possible, new respondents should actually use the full assessment for their scoring step as this will help increase their familiarity with the tools structure and the principles that the tool seeks to measure. Providing extra training and introduction to new respondents will enable consistency and standardization in their responses, ensuring more usable results.

Train Respondents on the Abridged Concept. All respondents should be briefly trained on the abridged assessment, to understand why practices are no longer presented in the same order and to understand why the second-phase practices vary for each respondent. Although the training does not necessarily cover the technical background of the prediction model that is used for selecting follow-up practices, it should make clear that the assessment uses data from a range of organizations to intelligently select potentially relevant follow-up questions. It is important that respondents have confidence in the prediction model and its ability to select useful follow-up practices. If needed, some of the graphs from the previous chapter can be used to reinforce the accuracy of the abridged LESAT and to help establish the prediction model as a credible and

useful approach. The training should emphasize both the advantages (efficiency, easier reassessment) as well as the disadvantages (less data granularity, inconsistency in the follow-up practices) of using an abridged assessment.

Determine Selection Criteria. For the full assessment, decision criteria for the evaluation of assessment results would occur during the fourth phase of the assessment process (see section 3.1.7). But for the abridged assessment, this step should be occur separately and earlier in the assessment, as the decision criteria used for evaluating results should be similar to the selection criteria that determines what follow-up practices appear in the second phase of the abridged assessment. For example, if the assessment users plan to act on the practices that have low current scores and large gaps, then that same combination should be used for selecting follow-up practices when respondents are actually completing the assessment.

6.2.2. Changes to Phase Three: Perform Assessment

In the second phase, the abridged LESAT requires changes to the assessment itself in order to handle the two-part assessment. Yet an important caveat to using the abridged LESAT is that discussion remains an important and valuable part of the process and must be retained, even if the assessment itself is performed electronically and, possibly, independently.

Conduct assessment. As demonstrated by the abridge assessment methodology presented previously (see Figure 9), the assessment is now a two-phase process. During the first phase, respondents score the six most indicative practices. These inputs are used to calculate predictions and determine the follow-up practices to present to the respondent. For the selection criteria used in the previous chapter, this meant twelve follow-up practices were presented and scored. As a result, the assessment step of the process must be changed to support the dynamic nature of the abridged assessment. The ideal approach is to perform the assessment digitally

enabling follow-up practices to be identified instantly and creating a seamless assessment process.[2] If preferred, paper assessment could still be retained, scoring the first six practices at the end of the introduction/training meeting providing time for the manual responses to be coded and follow-up practices to be selected. Each respondent would receive a personalized packet of twelve follow-up practices during the following meeting that would help in the identification of high-priority areas.

The disadvantage to electronic assessment is that respondents would not have the aid of a facilitator or an opportunity to ask clarification questions, as they would most likely be completing the assessment independently (this could be overcome using laptops or other mobile interfaces). On the other hand, paper assessment has the disadvantage of breaking the assessment process up into two distinct meetings, interrupting focus and potentially harming the quality of results.

Discuss and analyze results. Because respondents may not necessarily be together during the assessment process nor will they all see the same practices during the follow-up phase of the assessment, the step of discussion and analysis is even more important to the process. It allows respondents to gain clarification and exposure to high-priority practices that may not have appeared on their specific follow-up assessment.

Respondents should be presented with the data from the assessment process, but with the added dimension of the number of respondents per practice. More details on how to present and interpret results specifically for the abridged assessment appear in the next section (6.3).

In addition to the normally defined discussion that occurs during this step (see section 3.1.6), respondents should also discuss variations in the selection of follow-up practices.

[2] An online assessment tool was created in conjunction with this thesis: www.onlinelesat.com.

Although it may not be readily apparent why different follow-up practices were selected for different respondents, it is possible to compare and contract differences six most predictive practices to find clues regarding differences in the follow-up practices. For practices that have only one or two respondents, the group should aim to reach consensus on whether or not those practices are truly high-priority practices. Finally, the group should also briefly review those practices that do not appear in the results, as these may represent organizational strengths or may be indicative of high-priority areas not captures by the abridged assessment.

As an alternative process, an initial discussion can be held after scoring the first set of practices (the most indicative practices). During that discussion, the group can aim to reach consensus for all six practices, allowing a consistent set of follow-up practices to be scored by all respondents. The advantage of this is that consensus is built into the assessment process, creating more discussion and allowing for all respondents to be operating on a common-set of practices. The disadvantage is that the group no longer explores as diverse a range of high-priority practices, increasing the likelihood that certain viewpoints or certain potential weaknesses are not captured by the assessment.

6.2.3. Changes to Phase Four: Evaluate Assessment Results and Process

The fourth phase of the assessment process requires a few minor adjustments to both steps to reflect the unique nature of the abridged assessment tool.

Evaluate results to identify areas for improvement. Equipped with both the traditional interpretation tool, as well as those introduced in the next section (6.3), assessment users will perform the same evaluation of the results. Yet, since the abridged assessment captures one predominant subset of the results (as determined by the selection criteria used), practices cannot

be classified into distinct groups as with the normal evaluation method (since most practices will already be representative of one group). For example, if the selection criterion was "lowest current scores," then it would be hard to make a broad statement about practices with large gaps (since practices were specifically selected based on the lowest current score, the best that could be done is to identify those practices with both low current scores *and* large gaps). Because of this fact, all follow-up practices should be considered potential areas for enterprise improvement and should be reviewed accordingly, looking at such factors as respondent consensus and relative scoring differences. The practices used by the predictive model should also be considered as potential areas for improvement, but also serve as a measuring stick over multiple assessments for tracking the progress and success of past improvement efforts. In this way, the abridged assessment continues to fill two primary uses of assessment (identifying areas for improvement and tracking past improvement efforts).

Evaluate assessment process. In conjunction with the normal review of the assessment process, assessment users and respondents should discuss the two key factors of the abridged assessment. First, the group should review the selection criteria used to see if relevant follow-up practices were identified. By changing the selection criteria, the group may improve the assessment process in future cycles to better identify actionable high-priority practices. Second, the group should plan whether or not the abridged assessment should be used during the next assessment cycle. Although we have demonstrated that the abridged assessment is a valuable assessment tool, it only captures a portion of the data found in the full assessment. As a result, the enterprise should occasionally opt to repeat the full assessment process, especially at time of critical strategic planning (e.g., when drafting a new five year plan).

6.3. Interpreting the Results

The last section covered some of the key differences in the assessment process when using the abridged LESAT. Due to the distinct results from the abridged assessment, there are some additional interpretation methods for the results of the abridged assessment.

Selection of Follow-Up Practices. The first additional interpretation method available when using the abridged LESAT is to compare the selection of follow-up practices across respondents. Practices that appear in all respondents' follow-up section have greater consensus and are more likely to be high-priority practices for the enterprise, while practices that appear rarely or not at all are more likely to be enterprise strengths or of low strategic importance. Must like variance or intergroup discrepancies, follow-up selection can provide a lot of subtle insights. For example, if only a certain subgroup of respondents were asked a specific follow-up practice, it suggests that the practice is either more relevant or of more concern to that subgroup. By delving into the selection of follow-up practices, the abridged LESAT offers insights regarding group consensus and differences in priorities that may not be captured by simply variation in the scores (as there may not be sufficient responses to truly compare scores for a specific practice).

Trends in Practices. Since the abridged assessment enables more frequent reassessment, there is more opportunity for analysis of trends. As mentioned before, trends in the scores for the six most indicative practices provide insights into an overall measurement of enterprise performance. Similarly, trends in the follow-up practices identified may reinforce decisions about the most critical areas for improvement. For example, if one practice is repeatedly identified as a high-priority area during multiple reassessments, it should drive increased emphasis and focus on the specific practice.

Difference between Predictions and Previous Assessments. A final interpretation strategy for analyzing the abridged assessment results is to look at the predictions for those practices that do not appear in the selection of high-priority practices. The predicted scores can be compared to past full assessments in order to find discrepancies – that is, practices where the prediction seems to differ from the actual scores during the last full assessment. By reviewing these discrepancies, the assessment users may find one of two things. On one hand, the discrepancy may be indicative of improved performance with regards to the practice in question (that is, the enterprise truly improved performance and that is captured in the predictive model). On the other hand, the discrepancy may be due to a shortcoming of the predictive model reaching an inaccurate prediction. Either of these conclusions provides added insights to regarding improvements or other potential high-priority practices, and should be carefully considered during the interpretation step.

6.4. Potential Pitfalls

As with a standard full assessment, there are potential pitfalls associated with the abridged assessment that undermine the benefits of the process. In implanting the abridged version of LESAT, it is important that such pitfalls are avoided.

Not Performing Full Assessments. As has been stressed, the advantages of the abridged assessment are most suited to reassessment in order to follow-up on past insights, track progress and identify key next steps. Although it is tempting to use the abridged assessment from the beginning in order to reduce the time and resource overhead, it will overall hurt the assessment process. In addition, full assessments should be repeated on a semi regular basis (ideally, in sync with key strategic planning milestones). Not having a full LESAT as a reference point can harm the abridged assessment in several ways:

- Respondents are not familiar with the full range of the tool, and hence cannot judge the implications of the selected follow-up practices.
- Respondents may not trust the abridged tool, as it will feel overly targeted at certain practices. Doing comprehensive assessment will help build buy-in, commitment and trust for the assessment process by showing a range of principles all aligned towards enterprise excellence.
- There may be insufficient historical data to provide a baseline for interpreting the selection of follow-up practices.

Because of these risks, the enterprise should ensure that the full assessment is repeated every three to six years and all new respondents are at least exposed to all the practices in the full assessment to better understand the value of the process.

Interpreting Outside Selection Criteria. During the evaluation phase, it is easy to mistakenly interpret the results outside the range of the selection criteria. This can be seen most clearly if the selection criterion is only the lowest current score, yet the review of results interprets the highest current score. Because of the targeted nature of the abridged assessment, it specifically picks follow-up questions based on the predetermined selection criteria. This selection bias limits the nature of the data collected. Although the data is useful for answering questions regarding the selection criteria, it falls short when trying to answer independent questions. Even if the selection criterion lowest current scores, trying to answer questions about the largest gap will be unsuccessful. Instead, the abridged assessment can only answer questions to the extent that they overlap with the selection criteria. In the previous example, even though the abridged assessment would be unable to identify the largest gaps, it could provide

information on the practices with both a low current score *and* a large gap. This example is illustrated in Figure 14.

In order to avoid this pitfall, the review process and analysis must be cognizant of the selection criteria and the implications of the selection criteria on the choice of follow-up practices. Relative analysis within this scope is acceptable (ranking low current scores based on the size of the gap for example), but the insights and implications of such analysis must be properly scoped based on the initial selection criteria.

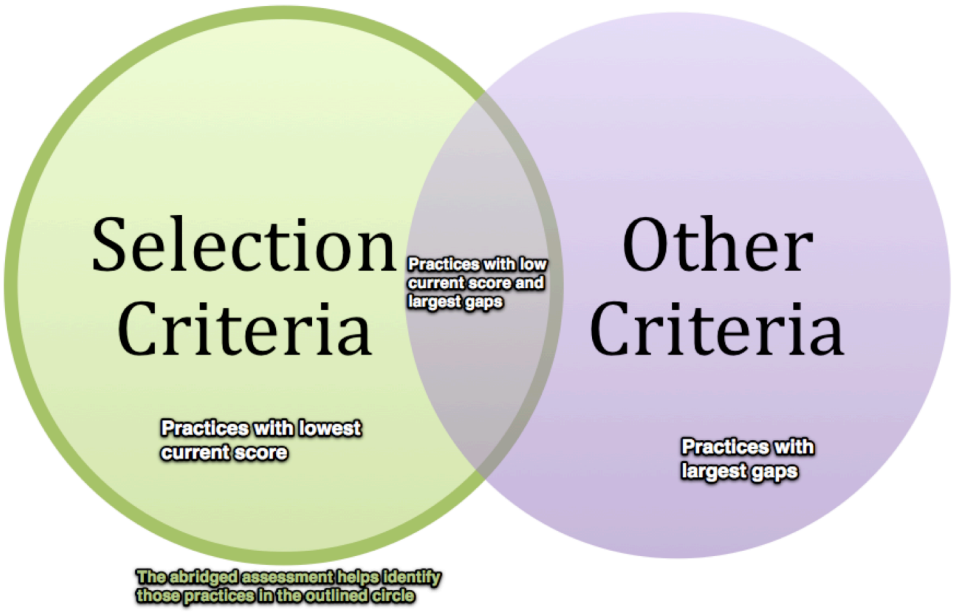


Figure 14 – Insights from the abridged assessment

Ignoring Practices without Scores. A final potential pitfall in the results stage of the abridged assessment is to ignore those practices that were not scored by any respondents. It is easy to ignore such practices, as they are not likely to be high-priority practices or to be crucial areas for enterprise improvement. Nonetheless, these practices still provide insights merely in their absence. For those practices that were not scored, the facilitators and users should briefly confirm that the practice is not a high-priority area, but they should also use such an insight as a

reference point. For example, the disappearance of a practice from the high-priority list may be indicative of successful improvement in a key area. Enterprise strengths should be both celebrated and recognized as strategic competencies that help differentiate the enterprise from its competition (Prahalad & Hamel, 1990). Although assessment helps in identifying areas for improvement, it should also reinforce the positive strengths and values of the enterprise, both as a tool to build morale and to reinforce competitive differentiation.

6.5. Summary

This chapter has reviewed the actual implementation of the abridged LESAT created in the previous chapter. By modifying the recommended LESAT practice to add additional planning and discussion steps, the abridged assessment can be easily deployed in a way that is both efficient and does not compromise the value of group discussion and consensus building. Once the abridged assessment is performed, standard interpretation tools can be complemented with insights about the distribution of follow-up practices. As long as the common pitfalls are avoided, the abridged assessment can be implemented in a way that measures overall enterprise performance and identifies key areas for improvement, while still highlighting existing enterprise strengths and competencies.

Chapter 7. Conclusions and Future Work

As the initial chapters established, enterprise assessment is a valuable and important tool for gaining and maintaining a competitive advantage in an era of global competition and innovation. Enterprise assessment enables enterprises to think and grow in a holistic manner, realizing potential areas for improvement that would normally be overlooked in local assessment and optimization. The increasing need for enterprise differentiation has led to literature on enterprise excellence (Rouse, 2005) as well as the construction of transformation frameworks to guide improvement initiatives at the enterprise level (Nightingale, 2009).

To accompany and aid in transformation planning, assessment tools have been developed to measure integrated and holistic indicators of enterprise performance (Clark & Kaehler, 2006; Shan, 2008). These tools pose numerous advantages, both direct (insights about performance, potential areas for improvement and organization/industry trends) as well as indirect (enabling enterprise decision making, encouraging consensus and providing greater transformation ownership). Despite these advantages, assessments remain underused due to high resource cost (Abdimomunova, Perkins, Valerdi, Shields, & Nightingale, 2011).

This thesis has set out to establish a methodology for creating abridged assessment tools that alleviate the time commitment required to assess an organization. This methodology was applied to the Lean Enterprise Self-Assessment Tool – an assessment with a large number of practices and a large intended group of respondents, making it an ideal assessment for the methodology (as time benefits will be multiplied across the whole group of respondents). In this final chapter, we will review the goals set out in the first chapter to see if the methodology and the abridged LESAT achieved the desired results.

7.1. Conclusions

In the introduction, we established three primary research hypotheses for this thesis. Based on the general methodology established and, specifically, the abridged LESAT, this section reflects on those hypotheses. To review, the three hypotheses are: (1) an abridged assessment can use a predictive model to identify, target and confirm high-priority areas for improvement; (2) the standard assessment process can incorporate the abridged assessment enabling enterprises to increase assessment frequency without proportionally increasing the resource burden; and (3) the abridged assessment can yield key insights related both to improving performance and potential areas for further improvement.

The first hypothesis was demonstrated and confirmed for the case of LESAT (Lean Enterprise Self-Assessment Tool). By applying a general methodology involving iterative feature selection and least squarest estimation, a predictive model was constructed that accurately predicted scores for other practices based on a set of six highly indicative practices. By combining the predictions with selection criteria, the predictive model allowed for the selection of follow-up practices that targeted high-priority areas (as defined by the selection criteria). When tested with actual assessment results, the predictive model was a significantly likely to identify relevant follow-up practices (on average, 93% of the desired high-priority practices were included in the follow-up scoring). This suggests that the abridged assessment is able to target a desired high-priority area, and use follow-up measurements to confirm this selection.

The abridged LESAT fits easily into the standard recommended assessment process confirming the second hypothesis. Only minor adjustments to the assessment process are required to implement the abridged assessment. The minor adjusts include: introducing the abridged assessment concept in training; deciding on selection criteria (what “type” of practices

should be targeted by the follow-up section); performing the assessment electronically while retaining the dialog and consensus building; and integrating additional tools for interpreting assessment results. With these adjustments to the assessment process in places, the abridged assessment can applied with much the same process as a traditional full assessment. Yet since the abridged assessment has only a fraction of the practice as the full assessment (in the case of LESAT, one third of the practices appear in the abridged assessment), there will be less time burden. The increased accessibility of assessment and the decreased time obligation required for the process allows enterprises to assess more regularly.

Finally, the third hypothesis is upheld by looking at the potential data that can be extracted from the abridged assessment process. Changes to the indicative practices provide ongoing measures of enterprise performance, since all respondents score these practices. The targeted, follow-up practices enable identification of key areas for improvement and provide the necessary inputs and guidance for transformation planning. Additionally, the abridged assessment offers other insights in terms of strengths (those practices that do not appear in or disappeared from the follow-up set of practices), group cohesion (consistency in the selection of follow-up practices across an organization) and outliers (by comparing predictions with past full assessments). Overall, the abridged assessment is data rich and provides many of the benefits of the full assessment.

These three hypotheses validate the abridged assessment that can both increase the appeal of regular assessment, yet still offer many of the benefits of full assessment. Although the process has solely been applied to one assessment in this thesis, it is generalizable to other enterprise assessment tools (and potentially to other assessment tools). By capitalizing on latent factors that link many practices assessed in such a tool, a predictive model can make inferences

about all scores based on an indicative subset. These predictions then allow targeted follow-up questions to reduce reliance on the prediction model. When applied to an assessment like LESAT, it results in a streamlined abridged tool that assess those practices that are most relevant and important in the strategic and transformation planning with less burden on the respondents. If assessment can be achieved with fewer demands on the time respondents – an important barrier to assessment currently – enterprises will be more likely established iterative assessment cycles with closed-loop feedback, allowing the assessment results to influence resource allocation and transformation prioritization. Overall, this closed-loop control structure will better enable enterprises to achieve improved performance (Hallam, 2003) and be more responsive to changing internal and external conditions.

7.2. Future Research

The work covered in the proceeding chapters examines how existing techniques such as feature selection and prediction models can be used in a novel methodology to decrease the number of assessment inputs for enterprise assessment. As a result, there are ample opportunities for ongoing research and work in this domain. The methodology was validated when applied to LESAT, as an abridged LESAT was created with only one third of the practices. Yet the training and testing data for the abridged LESAT all came from medium to large aerospace companies in the United States and the United Kingdom. To truly confirm the accuracy and usability of the abridged assessment, it should be trained and tested in more sectors. In addition, the abridged LESAT seeks to make predictions and inferences on a respondent-by-respondent basis. Yet each respondent has a larger enterprise context that may contribute to increased prediction accuracy. By looking at respondents in their enterprise context, accuracy may be improved and additional predictions about group cohesion may be possible. Looking beyond LESAT, the methodology

can be further validated by applying it to other enterprise assessment tools. A final area for further research is the piloting of the abridged assessment tool. Although the process has been used once, more data would confirm whether the process is intuitive, useful and easy to execute. Performing this additional research will help refine abridged assessment tools and increase the feasibility of regular assessment to enable effective enterprise transformation.

References

Abdel-Khalek, A. M. (2006). Measuring Happiness With a Single-Item Scale. *Social Behavior and Personality: an international journal*, *34*(2), 139-150. doi: 10.2224/sbp.2006.34.2.139.

Abdimomunova, L. (2010). *Organizational Assessment Processes for Enterprise Transformaton*. Massachusetts Institute of Technology.

Abdimomunova, L., Perkins, L. N., Valerdi, R., Shields, T., & Nightingale, D. J. (2011). Enterprise Assessment Process. *Information, Knowledge, Systems Management*. Cambridge.

Anderson, J. C., Gerbing, D. W., & Hunter, J. E. (1987). On the Assessment of Unidimensional Measurement: Internal and External Consistency, and Overall Consistency Criteria. *Journal of Marketing Research*, *24*(4), 432. doi: 10.2307/3151392.

ASQC. (2009). *Baldrige National Quality Program: Criteria for Performance Excellence*. Milwaukee, WI: ASQC Quality Press.

Bourne, M., Franco, M., & Wilkes, J. (2003). Corporate performance management. *Measuring Business Excellence*, *7*(3), 15-21. doi: 10.1108/13683040310496462.

Burton, R. M., Lauridsen, J., & Obel, B. (2004). The impact of organizational climate and strategic fit on firm performance. *Human Resource Management*, *43*(1), 67-82. doi: 10.1002/hrm.20003.

Clark, D. L., & Kaehler, T. J. (2006). *Evaluation of Organizational Self-Assessment Tools and Methodologies to Measure Continous Improvement for the Naval Aviation Enterprise*. *Applied Sciences*. Naval Postgraduate School.

Collins, J. (2001). *Good to Great: Why Some Companies Make the Leap... and Others Don't*. New York: Harper Business.

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of applied psychology*, *78*(1), 98-104. APA AMERICAN PSYCHOLOGICAL ASSOCIATION.

Creswell, J. W. (2002). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches* (2nd ed.). Thousand Oaks: Sage Publications.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*(3), 297-334.

De Ven, A. H. van, & Ferry, D. L. (1980). *Measuring and Assessing Organizations* (p. 552). New York: Wiley.

Dellinger, A. B., & Leech, N. L. (2007). Toward a Unified Validation Framework in Mixed Methods Research. *Journal of Mixed Methods Research*, *1*(4), 309-332.

Denison, D. R., & Mishra, A. K. (1995). Toward a Theory of Organizational Culture and Effectiveness. *Organization Science*, *6*(2), 204223. INFORMS.

Donaldson, S. I., & Grant-Vallone, E. J. (2002). Understanding self-report bias in organizational behavior research. *Journal of Business and Psychology*, *17*(2), 245–260. Springer.

Furr, M. R., & Bacharach, V. R. (2008). *Psychometrics: An Introduction*. Thousand Oaks: Sage Publications.

Gardner, P. L. (1995). Measuring attitudes to science: Unidimensionality and internal consistency revisited. *Research in Science Education*, *25*(3), 283-289. doi: 10.1007/BF02357402.

George, W. (2006). Transformation Leadership. In W. B. Rouse (Ed.), *Enterprise Transformation* (pp. 69-78). Hoboken, NJ: John Wiley and Sons.

Godard, J. (2004). A Critical Assessment of the High-Performance Paradigm. *British Journal of Industrial Relations*, *42*(2), 349-378. doi: 10.1111/j.1467-8543.2004.00318.x.

Guyon, I., & Elisseeff, A. (2003). An Introduction to Variable and Feature Selection. *The Journal of Machine Learning Research*, *3*, 1157–1182. JMLR. org.

Hallam, C. R. A. (2003). *Lean Enterprise Self-Assessment as a Leading Indicator for Accelerating Transformation in the Aerospace Industry*. MIT.

Hallam, C. R. A. (2004). Enterprise Transformation: Measures and Incentives. *Aerospace*. Northgrop Gruman Integrated Systems: Lean Aerospace Initiative Plenary Conference.

Hauser, J. R., & Katz, G. M. (1998). Metrics: you are what you measure! *European Management Journal*, *16*(5), 517–528. Elsevier.

Hoerl, R. W. (1998). Six Sigma and the future of the quality profession. *Quality Progress*, *31*(6), 35-39. THE IEEE, INC.

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

Jolliffe, I. T. (2002). *Principal Component Analysis* (2nd ed., p. 518). New York: Springer-Verlang.

Jost, J. T., Banaji, M. R., & Nosek, B. A. (2004). A Decade of System Justification Theory: Accumulated Evidence of Conscious and Unconscious Bolstering of the Status Quo. *Political Psychology*, *25*(6), 881-919. doi: 10.1111/j.1467-9221.2004.00402.x.

Kotnour, T. (2011). An Emerging Theory of Enterprise Transformations. *Journal of Enterprise Transformation*, *1*(1), 48-70. doi: 10.1080/19488289.2010.550669.

Kotter, J. P. (1995). Leading change: Why transformation efforts fail. *Harvard Business Review*, *73*(2), 59-67.

Kueng, P. (2000). Process performance measurement system: a tool to support process-based organizations. *Total Quality Management*, *11*(1), 67–85. London: Routledge.

Lansing, E., & Si, L. (2004). A Study of Methods for Normalizing User Ratings in Collaborative Filtering. *Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval* (p. 2). New York: ACM Press.

LeBreton, J. M., & Senter, J. L. (2007). Answers to 20 Questions About Interrater Reliability and Interrater Agreement. *Organizational Research Methods*, *11*(4), 815-852. doi: 10.1177/1094428106296642.

Link, A., & Scott, J. (2006). An economic evaluation of the Baldrige National Quality Program. *Economics of Innovation and New Technology*, *15*(1), 83-100. doi: 10.1080/1043859042000332204.

Maskell, B. H. (1991). *Performance Measurement for World Class Manufacturing: A Model for American Companies*. New York: Productivity Press.

Miller, B. M. G. (2006). Thinking About Second & Third Order Effects: A Sample (And Simple) Methodology. *Defense*, 52-55.

MIT. (2000). *Transitioning to a lean enterprise: a guide for leaders — Volume I: Executive overview*. (D. J. Nightingale, Ed.). Cambridge: MIT Lean Aerospace Initative.

MIT. (2001). *Lean Enterprise Self-Assessment Tool Facilitator's Guide - Version 1.0*. Cambridge, MA: MIT Lean Aerospace Initative.

Montoya, M., Schofield, D. M., Chowdhury, A. A., & Lehman, R. J. (2009). *Lean Analysis for a Startup Health Care Engineering Firm*. Cambridge.

Moore, A. W. (2006). Cross-validation for detecting and preventing overfitting. *Science*. Retrieved January 11, 2009, from http://www.autonlab.org/tutorials/.

Murman, E., Allen, T., Bozdogan, K., Cutcher-Gershenfeld, J., McManus, H., Nightingale, D. J., et al. (2002). *Lean Enterprise Value Insights from MIT's Lean Aerospace Initiative*. New York: PALGRAVE.

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, *2*(2), 175-220. doi: 10.1037/1089-2680.2.2.175.

Nightingale, D. J. (2009). Principles of Enterprise Systems. Cambridge, MA: ESD Symposium.

Nightingale, D. J., & Mize, J. H. (2002). Development of a lean enterprise transformation maturity model. *Information, Knowledge, Systems Management*, *3*(1), 15–30. IOS Press.

Nightingale, D. J., & Srinivasan, J. (2011). *Beyond the Lean Revolution*. New York: McGraw-Hill.

Perkins, L. N., Abdimomunova, L., Valerdi, R., Shields, T., & Nightingale, D. J. (2011). Insights from Enterprise Assessment. *Information, Knowledge, Systems Management*.

Prahalad, C. K., & Hamel, G. (1990). The Core Competence of the Corporation. *Harvard Business Review*, *68*(3), 79-91.

Roth, G. (1999). Cracking the "black box" of a learning initiative assessment. *The Dance of Change* (p. 303–311).

Rouse, W. B. (2005). A theory of enterprise transformation. *Systems Engineering*, *8*(4), 279-295.

Rouse, W. B. (2011). Necessary Competencies for Transforming an Enterprise. *Journal of Enterprise Transformation*, *1*(1), 71-92. doi: 10.1080/19488289.2010.548905.

Shan, R. S. (2008). *The Role of Assessment in a Lean Transformation*. *Management*. Massachusetts Institute of Technology.

Sharp, C., Goodyer, I. M., & Croudace, T. J. (2006). The Short Mood and Feelings Questionnaire (SMFQ): a unidimensional item response theory and categorical data factor analysis of self-report ratings from a community sample of 7-through 11-year-old children. *Journal of abnormal child psychology*, *34*(3), 379-91. doi: 10.1007/s10802-006-9027-x.

Shingo Prize for Operational Excellence, T. (2009). *The Shingo Prize for Operational Excellence*. Logan, UT: Utah State University, Jon M. Huntsman School of Business.

Siviy, J. M., Penn, M. L., & Stoddard, R. W. (2007). *CMMI and Six Sigma: Partners in Process Improvement*. Boston: Addison-Wesley Professional.

Somol, P., Novovičová, J., & Pudil, P. (2010). Efficient Feature Subset Selection and Subset Size Optimization. In H. Adam (Ed.), *Pattern Recognition Recent Advances* (pp. 1-24). Croatia: InTech.

Stoppiglia, H., Dreyfus, G., Dubois, R., & Oussar, Y. (2003). Ranking a Random Feature for Variable and Feature Selection. (I. Guyon & Andre Elisseeff, Eds.)*Journal of Machine Learning Research*, *3*(7-8), 1399-1414. doi: 10.1162/153244303322753733.

Venkatraman, N. (1989). Strategic Orientation of Business Enterprise: The Construct, Dimensionality and Measurement. *Management Science*, *35*(8), 942-962. doi: 10.1287/mnsc.35.8.942.

Womack, J. P., Jones, D. T., & Roos, D. (1990). *The Machine that Changed the World: The Story of Lean Production*. New York: Harper.

Appendices

Appendix A: LESAT Structure

The table below shows the 54 LESAT practices and the respective organization of practices into sections and subsections.

Section I: Lean Transformation/Leadership
Subsection I.A: Enterprise Strategic Planning
I.A.1 Integration of Lean in Strategic Planning Process
I.A.2 Focus on Customer Value
I.A.3 Leveraging the Extended Enterprise
Subsection I.B: Adopt Lean Paradigm
I.B.1 Learning and Education in "Lean" for Enterprise Leadership
I.B.2 Senior Management Commitment
I.B.3 Lean Enterprise Vision
I.B.4 A Sense of Urgency
Subsection I.C: Focus on the Value Stream
I.C.1 Understanding the Current Value Stream
I.C.2 Enterprise Flow
I.C.3 Designing the Future Value Stream
I.C.4 Performance Measures
Part I.D: Develop Lean Structure & Behavior
I.D.1 Enterprise Organizational Orientation
I.D.2 Relationships Based on Mutual Trust
I.D.3 Open and Timely Communications
I.D.4 Employee Empowerment
I.D.5 Incentive Alignment
I.D.6 Innovation Encouragement
I.D.7 Lean Change Agents
Subsection I.E: Create & Refine Transformation Plan
I.E.1 Enterprise-Level Lean Transformation Plan
I.E.2 Commit Resources for Lean Improvements
I.E.3 Provide Education and Training
Subsection I.F: Implement Lean Initiatives
I.F.1 Development of Detailed Plans Based on Enterprise Plan
I.F.2 Tracking Detailed Implementation
Subsection I.G: Focus on Continuous Improvement
I.G.1 Structured Continuous Improvement Processes
I.G.2 Monitoring Lean Progress
I.G.3 Nurturing the Process
I.G.4 Capturing Lessons Learned

		I.G.5 Impacting Enterprise Strategic Planning
Section II: Life Cycle Processes		
	Subsection II.A: Business Acquisition and Program Management	
		II.A.1 Leverage Lean Capability for Business Growth
		II.A.2 Optimize the Capability and Utilization of Assets (People, equipment, facilities, etc.)
		II.A.3 Provide Capability to Manage Risk, Cost, Schedule and Performance
		II.A.4 Allocate Resources for Program Development Efforts
	Subsection II.B: Requirements Definition	
		II.B.1 Establish a Requirement Definition Process to Optimize Lifecycle Value
		II.B.2 Utilize Data from the Extended Enterprise to Optimize Future Requirement Definitions
	Subsection II.C: Develop Product and Process	
		II.C.1 Incorporate Customer Value into Design of Products and Processes
		II.C.2 Incorporate Downstream Stakeholder Values (Manufacturing, Support, etc.) into Products and Processes
		II.C.3 Integrate Product and Process Development
	Subsection II.D: Manage Supply Chain	
		II.D.1 Define and Develop Supplier Network
		II.D.2 Optimize Network-Wide Performance
		II.D.3 Foster Innovation and Knowledge-Sharing Throughout the Supplier Network
	Subsection II.E: Produce Product	
		II.E.1 Utilize Production Knowledge and Capabilities for Competitive Advantage
		II.E.2 Establish and Maintain a Lean Production System
	Subsection II.F: Distribute and Service Product	
		II.F.1 Align Sales and Marketing to Production
		II.F.2 Distribute Product in Lean Fashion
		II.F.3 Enhance Value of Delivered Products and Services to Customers and the Enterprise
		II.F.4 Provide Post Delivery Service, Support and Sustainability
Section III: Enabling Infrastructure		
	Subsection III.A: Lean Organizational Enablers	
		III.A.1 Financial System Supports Lean Transformation
		III.A.2 Enterprise Stakeholders Pull Required Financial Information
		III.A.3 Promulgate the Learning Organization
		III.A.4 Enable the Lean Enterprise with Information Systems and Tools
		III.A.5 Integration of Environmental Protection, Health and Safety into the Business
	Subsection III.B: Lean Process Enablers	
		III.B.1 Process Standardization
		III.B.2 Common Tools and Systems
		III.B.3 Variation Reduction

Appendix B: MSE at Each Training Iteration

When building the predictive model, each assessment input is considered as a potential new feature and the reduction in mean square error (MSE) for the predictive model is calculated. The below table illustrates this by showing each feature considered. Highlighted inputs represent the selected features – those that most reduce the MSE of the predictive model.

	Current			Desired			Gap		
Iteration	1	2	3	1	2	3	1	2	3
I.A.1	0.750	0.720	0.696	0.752	0.655	0.619	0.855	0.758	0.730
I.A.2	0.758	0.721	0.700	0.746	0.645	0.607	0.809	0.746	0.720
I.A.3	0.757	0.715	0.698	0.719	0.635	0.607	0.789	0.744	0.718
I.B.1	0.761	0.730	0.699	0.726	0.651	0.615	0.817	0.752	0.726
I.B.2	0.756	0.724	0.696	0.730	0.646	0.615	0.821	0.747	0.723
I.B.3	0.743	0.719	0.694	0.685	0.637	0.608	0.793	0.748	0.727
I.B.4	0.758	0.727	0.701	0.705	0.645	0.612	0.816	0.748	0.727
I.C.1	0.748	0.724	0.700	0.674	0.634	0.607	0.777	0.734	0.725
I.C.2	0.762	0.726	0.705	0.690	0.632	0.607	0.793	0.741	0.728
I.C.3	0.747	0.722	0.698	0.679	0.636	0.607	0.767	***0.729***	
I.C.4	0.744	0.720	0.699	0.713	0.639	0.604	0.820	0.744	0.729
I.D.1	0.740	***0.702***		0.704	0.633	0.610	0.799	0.751	0.722
I.D.2	0.748	0.713	0.698	0.712	0.642	0.611	0.801	0.742	0.717
I.D.3	0.761	0.721	0.700	0.734	0.651	0.611	0.801	0.748	0.720
I.D.4	0.760	0.710	0.696	0.716	0.643	0.609	0.813	0.746	0.716
I.D.5	0.748	0.721	0.697	0.717	0.649	0.615	0.796	0.741	0.720
I.D.6	0.757	0.719	0.699	0.697	0.639	0.612	0.789	0.736	0.713
I.D.7	0.758	0.727	0.701	0.683	0.641	0.613	0.804	0.743	0.724
I.E.1	0.750	0.729	0.705	0.663	0.640	0.610	0.777	0.739	0.720
I.E.2	0.751	0.721	0.698	0.720	0.651	0.616	0.832	0.761	0.729
I.E.3	0.752	0.727	0.700	0.674	0.642	0.607	0.780	0.737	0.718
I.F.1	0.738	0.723	0.697	***0.657***			0.783	0.739	0.724
I.F.2	0.747	0.729	0.702	0.677	0.645	0.611	0.789	0.744	0.726
I.G.1	0.749	0.728	0.702	0.703	0.649	0.617	0.795	0.744	0.721
I.G.2	***0.732***			0.668	0.636	0.616	0.791	0.739	0.719
I.G.3	0.743	0.720	0.697	0.721	0.653	0.618	0.807	0.747	0.719
I.G.4	0.766	0.727	0.702	0.694	0.638	0.609	0.787	0.746	0.721
I.G.5	0.736	0.717	0.697	0.695	0.633	0.611	0.805	0.744	0.720
II.A.1	0.745	0.727	0.701	0.674	0.641	0.614	0.795	0.748	0.721
II.A.2	0.774	0.727	0.701	0.700	0.640	0.613	0.798	0.743	0.721
II.A.3	0.749	0.711	***0.688***	0.713	0.640	0.613	0.822	0.759	0.729
II.A.4	0.746	0.710	0.692	0.708	0.639	0.610	0.802	0.746	0.720
II.B.1	0.752	0.714	0.697	0.674	0.625	0.611	0.785	0.754	0.727
II.B.2	0.753	0.718	0.695	0.673	0.622	0.607	0.785	0.749	0.716
II.C.1	0.755	0.712	0.689	0.735	0.631	0.607	0.808	0.751	0.717
II.C.2	0.756	0.713	0.691	0.719	0.624	0.611	0.811	0.757	0.727
II.C.3	0.746	0.707	0.693	0.705	0.620	0.610	0.781	0.747	0.718
II.D.1	0.745	0.709	0.690	0.690	0.627	0.606	0.770	0.746	0.721
II.D.2	0.753	0.714	0.691	0.700	0.626	***0.603***	0.784	0.746	0.717

II.D.3	0.756	0.722	0.698	0.708	0.650	0.632	0.770	0.745	0.721
II.E.1	0.759	0.715	0.693	0.741	0.664	0.638	***0.766***		
II.E.2	0.754	0.717	0.696	0.700	0.629	0.607	0.798	0.742	0.718
II.F.1	0.768	0.724	0.699	0.735	0.629	0.612	0.810	0.752	0.721
II.F.2	0.765	0.722	0.698	0.734	0.623	0.613	0.806	0.745	0.713
II.F.3	0.747	0.707	0.690	0.694	***0.618***		0.773	0.742	0.713
II.F.4	0.752	0.713	0.692	0.714	0.622	0.613	0.781	0.734	***0.709***
III.A.1	0.763	0.723	0.699	0.687	0.625	0.606	0.786	0.751	0.725
III.A.2	0.761	0.719	0.698	0.703	0.635	0.613	0.799	0.750	0.720
III.A.3	0.751	0.709	0.694	0.698	0.637	0.612	0.769	0.736	0.714
III.A.4	0.753	0.710	0.692	0.716	0.633	0.609	0.800	0.752	0.724
III.A.5	0.758	0.714	0.691	0.756	0.645	0.610	0.811	0.750	0.719
III.B.1	0.747	0.711	0.692	0.728	0.635	0.612	0.794	0.745	0.722
III.B.2	0.742	0.710	0.693	0.719	0.633	0.611	0.797	0.742	0.718
III.B.3	0.746	0.715	0.695	0.683	0.623	0.610	0.795	0.752	0.723

Appendix C: MSE for Final Predictive Model

Once the predictive practices were combined into a comprehensive model, the mean square error (MSE) of remaining practices was calculated again and is represented in the table below. Only practices not included in the predictive model are listed.

Practice	Current	Desired	Gap
I.A.1 Integration of Lean in Strategic Planning Process	0.2495	0.1397	0.2147
I.A.2 Focus on Customer Value	0.2206	0.2409	0.2357
I.A.3 Leveraging the Extended Enterprise	0.2093	0.1843	0.2002
I.B.1 Learning and Education in "Lean" for Enterprise Leadership	0.1427	0.2268	0.2245
I.B.2 Senior Management Commitment	0.2246	0.2631	0.3612
I.B.3 Lean Enterprise Vision	0.1387	0.1695	0.2463
I.B.4 A Sense of Urgency	0.1299	0.2528	0.2586
I.C.1 Understanding the Current Value Stream	0.1716	0.179	0.1812
I.C.2 Enterprise Flow	0.1324	0.1517	0.1586
I.C.4 Performance Measures	0.1976	0.1889	0.272
I.D.2 Relationships Based on Mutual Trust	0.1774	0.1846	0.2197
I.D.3 Open and Timely Communications	0.1663	0.15	0.2132
I.D.4 Employee Empowerment	0.3065	0.1802	0.2301
I.D.5 Incentive Alignment	0.1576	0.2606	0.2772
I.D.6 Innovation Encouragement	0.1923	0.156	0.2168
I.D.7 Lean Change Agents	0.2008	0.1783	0.2373
I.E.1 Enterprise-Level Lean Transformation Plan	0.1296	0.1281	0.204
I.E.2 Commit Resources for Lean Improvements	0.1584	0.1703	0.2426
I.E.3 Provide Education and Training	0.1303	0.1457	0.2321
I.F.2 Tracking Detailed Implementation	0.1609	0.1397	0.1637
I.G.1 Structured Continuous Improvement Processes	0.1516	0.1516	0.172
I.G.3 Nurturing the Process	0.1723	0.1431	0.2129
I.G.4 Capturing Lessons Learned	0.1436	0.1462	0.1634
I.G.5 Impacting Enterprise Strategic Planning	0.2072	0.1848	0.2422
II.A.1 Leverage Lean Capability for Business Growth	0.1217	0.128	0.1807
II.A.2 Optimize the Capability and Utilization of Assets (People,	0.147	0.1599	0.2037

equipment, facilities, etc.)			
II.A.3 Provide Capability to Manage Risk, Cost, Schedule and Performance	0.3574	0.1696	0.2065
II.A.4 Allocate Resources for Program Development Efforts	0.2273	0.1442	0.1876
II.B.1 Establish a Requirement Definition Process to Optimize Lifecycle Value	0.1845	0.1675	0.1932
II.B.2 Utilize Data from the Extended Enterprise to Optimize Future Requirement Definitions	0.2081	0.2394	0.1964
II.C.1 Incorporate Customer Value into Design of Products and Processes	0.2345	0.2859	0.2374
II.C.2 Incorporate Downstream Stakeholder Values (Manufacturing, Support, etc.) into Products and Processes	0.1883	0.1737	0.2462
II.C.3 Integrate Product and Process Development	0.1866	0.2396	0.1856
II.D.1 Define and Develop Supplier Network	0.1712	0.1428	0.2198
II.D.2 Optimize Network-Wide Performance	0.1532	0.1528	0.2038
II.D.3 Foster Innovation and Knowledge-Sharing Throughout the Supplier Network	0.1252	0.1996	0.1749
II.E.2 Establish and Maintain a Lean Production System	0.2007	0.2331	0.2172
II.F.1 Align Sales and Marketing to Production	0.2644	0.2538	0.2884
II.F.2 Distribute Product in Lean Fashion	0.1941	0.2871	0.3208
II.F.4 Provide Post Delivery Service, Support and Sustainability	0.2678	0.2476	0.2343
III.A.1 Financial System Supports Lean Transformation	0.1729	0.2048	0.2276
III.A.2 Enterprise Stakeholders Pull Required Financial Information	0.1419	0.1795	0.2486
III.A.3 Promulgate the Learning Organization	0.2368	0.194	0.2076
III.A.4 Enable the Lean Enterprise with Information Systems and Tools	0.3065	0.1793	0.2128
III.A.5 Integration of Environmental Protection, Health and Safety into the Business	0.2906	0.2624	0.2692
III.B.1 Process Standardization	0.2644	0.1449	0.2108
III.B.2 Common Tools and Systems	0.2328	0.171	0.2034
III.B.3 Variation Reduction	0.1615	0.2053	0.22
Average	0.1938	0.189	0.2223