Conceptual Framework for adapting Technical Performance Measurement Methodology for Early Stage Research and Development Projects

by Ross E. Agee

B.S. in Biochemistry, May 2006, University of Nebraska-Lincoln
M.S. in Project Management, May 2013, The George Washington University

A Praxis submitted to

The Faculty of
The School of Engineering and Applied Science
of The George Washington University
in partial satisfaction of the requirements
for the degree of Doctor of Engineering

January 19, 2018

Praxis directed by

Justin Eggstaff
Professorial Lecturer of Engineering Management and System Engineering

Muhammad Faysal Islam
Professorial Lecturer of Engineering Management and System Engineering

The School of Engineering and Applied Science of The George Washington University certifies that Ross Eugene Agee has passed the Final Examination for the degree of Doctor of Engineering as of August 26, 2017. This is the final and approved form of the praxis.

Conceptual Framework for adapting Technical Performance Measurement Methodology for Early Stage Research and Development Projects

Ross E. Agee

Praxis Research Committee:

Justin Eggstaff, Professorial Lecturer of Engineering Management and System Engineering, Praxis Co-Director

Muhammed Islam, Professorial Lecturer of Engineering Management and System Engineering, Praxis Co-Director

O. Erik Strack, Manager, Scalable Algorithms, Sandia National Laboratories, Committee Member

Shahram Sarkani, Professor of Engineering Management and Systems Engineering, Committee Member

Thomas Andrew Mazzuchi, Professor of Engineering Management and Systems Engineering & of Decision Sciences, Committee Member

Acknowledgement

There is no amount of text or pages that can convey all the appreciation that is deserved for those who helped me along the path to completing this praxis. This page is a meager attempt to convey volumes of thanks to all those who deserve it.

First, my beautiful wife-to-be Jenny Lee Hansen. Having put up with my sour attitude, stress, and distracted mind over the course of this effort could not have been easy. Thank you much for your encouragement, understanding, and counsel.

Next, Dr. O. Erik Strack is owed many, many, many thanks. I find it very fitting that your words served as bookends on this journey. You provided encouragement and guidance in the beginning, then critical feedback at the end. I cannot thank you enough for helping me and improving this paper.

I appreciate all of my family and soon-to-be family standing by during this process. I wasn't always available or present during our functions while my mind or mind and body were working on this paper. You never batted an eye and understood fully. I thank you for being so understanding and encouraging.

I thank my coworkers for putting up with me during this time. I couldn't help but have my mind wander away from topics into thinking about this paper. Thanks for understanding and picking up the slack when I was absent.

Thank you to Dr. Barry H. Rabin, Joe Cleary, Jeff Chamberlin, and Scott Ravenhill for providing the data for this research. It would literally have been impossible without you.

Finally, thank you to Dr. Justin Eggstaff and Dr. Muhammed Islam for your guidance throughout this process. Your time and dedication is very much appreciated.

Abstract

Conceptual Framework for adapting Technical Performance Measurement Methodology for Early Stage Research and Development Projects

Early stage research and development (R&D) projects, programs, or efforts are characterized by a high degree of uncertainty. In the public sector, these efforts are defined as those with technology readiness levels up to 4. Both project management and systems engineering cultures proliferate throughout the public sector. Project management traditionally uses earned value management system to track the progress of its projects. Systems engineers apply Technical Performance Measurement methodology (TPM) to R&D efforts to monitor progress of quality factors with respect to those efforts. Both of these methods have numerous alternative implementations, and both strive to provide risk management indicators and insight into future performance. This paper discusses a conceptual framework for monitoring early stage R&D by altering a TPM method. By quantifying and monitoring the "learning" or technical uncertainty toward the project goal, early stage R&D is given a measurement for technical progress. Learning values are determined using pairwise comparisons of R&D activities that contribute to learning. How the effort increases learning, or degrades uncertainty, over time serves as the baseline for performance management. In order to predict future performance, the learning baseline is regressed in linear segments. Using simulated project status, those predictions are shown to have acceptable uncertainty by validating against prediction intervals.

Table of Contents

List of Figures

List of Tables

1 Introduction

1.1 Research and Development Management

Innovation is an everyday buzzword in many large organizations. The need to innovate and stay ahead of the market is an essential element in maintaining a competitive edge (Dustin, Gentet, & Jitendra, 2014). Consistent with any process in the workplace, there is an inherent need to manage innovation.

Commercial organizations their management processes and procedures that value the commercial necessity of innovation – the potential profits (Adams, Bessant, and Phelps, 2006). These management processes are tailored to monitor inputs and output potential from the perspective of financial return (Kahn & McGourty, 2009). Within that management, commercial organizations inherently manage the expected utility or effectiveness of the new product balanced against investment and profitability (Adams et al., 2006). Many functions can assist innovation management and apply their inputs including marketing, engineering, and finance (Schilling, 2013, p. 263).

Public sector organizations are not competing in the commercial market. They are concerned with R&D that is difficult to assign a market value, although they do need to balance their investments (Eckhause, Hughes, & Gabriel, 2009). Within that paradigm, public sector organizations manage innovation to meet capabilities, solve problems, and perform functions while concerning themselves with the amount of money being invested.

Systems engineering and project management cultures are strongholds in the public sector. Many organizations have adopted project management cultures, and emphasized the tools and techniques of that system (Crawford & Helms, 2009). Systems engineering has long been applied to acquire major systems (Locatelli, Mancini, & Romano, 2014). Project Management Professional and Certified Systems Engineering Professional certifications are in vogue.

As modern communication systems and technology have increased in capabilities, business analytics have risen as a priority, with 83% of Chief Information Officers stating it was their number one priority in 2009 and 2011 (Holsapple, 2014). Considerable attention is devoted to gathering data and understanding its relationship to performance (Davenport, 2013). Business leaders hope to utilize that data to predict future outcomes and extrapolate meaningful information about their business operations (Davenport, 2013).

Using data to assess performance, like business analytics, has been an inherent function of project management for many years. Since its adoption by the federal government in 1967, Earned Value Management System (EVMS) metrics have been reported against projects to inform project managers about schedule, cost, and work performed (Kwak & Anbari, 2012). Iterations and alterations to EVMS have been developed and tried across the discipline including forecasting methodologies (Vandevoorde & Vanhoucke, 2006).

The Department of Defense started implementing project management in the early 1960s and helped develop modern project management tools. The National Air and Space Administration also implemented these project management tools in 1997 (Kwak & Anbari, 2012). Currently, the government directs EVMS procedures for every project totaling over $20 million (Johnson, 2006).

The public sector has adopted project management and integrated it in its culture (Crawford & Helms, 2009). As innovation continues to be a priority, the public sector continues to pursue and fund innovation. Since these can be treated as projects, the project management culture may seep into these innovation or research and development projects. Project managers may utilize the emerging project tools and management metrics to track these research and development projects.

Project management, and more specifically EVMS, tend to focus on accomplishing the project within cost, scope, and schedule (the triple constraint) (Meredith & Mantel). However, there is a unique difference in project success and project management success (Locatelli, Mancini, & Romano, 2014). It is possible to complete a project within the triple constraint and not meet the needs of the customer. Ika (2009) expanded on this thought by stating:

> In our journey toward a comprehensive understanding of project success, one should not confuse any more between project management success and project success. Semantically, project management success refers to efficiency, an internal concern to the project team, and project success embraces concerns for efficiency and effectiveness—in other words, all concerns, whether internal or external, short-term or long-term.

Systems engineering can bridge the gap between project management success and project success, as it requires success to be defined in the context of utility and not just the triple constraint. According to the International Council on Systems Engineering, "Systems Engineering considers both the business and the technical needs of all customers with the goal of providing a quality product that meets the user needs" (INCOSE, 2017). Systems Engineering developed in the US following World War II, has evolved into a conglomeration of multiple disciplines concerned with the management of the entire lifecycle of a system (Locatelli, Mancini, & Romano, 2014).

Systems engineering can use technical performance measurements and their associated methods as tools to assess the technical progress of research and development activities. Generally, technical performance measures are measurable and quantified requirements or desires for system performance (Roedler & Jones, 2005). The progress of the achievement of these performance measures is tracked and measured as the system goes along in its development (Sears & Taylor, 1984). These tools can serve to bridge the gap between project management success and project success by giving managers the methods to track the technical performance within the good customer-value based systems engineering principles (DAU, 2017a).

1.2 Research and Development Considerations

Research and development can be used as an overarching term encompassing "work directed on a large scale towards the innovation, introduction, and improvement of products and processes" (Oxford English Dictionary, n.d.). Innovation, new product development, science and technology, and radical and incremental innovation are all terms that could fall within this definition. In this use, the terminology covers a broad swath of technology evolution and is mostly generic in its description. Numerous models exist to describe the evolution of innovation (Godin, 2015).

For the context of the following research, research and development will be defined within the three phases of basic science, applied research, and product development. In the beginning of research and development, the concepts are formed, and in the end, a fully realized solution is brought to fruition. It is important to understand and define the components of research and development, in order to focus the subsequent discussion. It can be segmented into these three major areas (Privitera, Design, & Johnson, 2009).

1.2.1 Basic Science

At the inception of research and development is basic science (Godin, 2006). In this phase, the underpinning science that enables research and development is established. The important distinction from other components of research and development is that basic science does not necessarily presuppose an application (Defense Science Board, 2012). Basic science answers fundamental questions about materials, compatibilities, processes, and interactions (NSF, 1953). Since this is fundamentally a knowledge growth phase, there is a high amount of learning to be achieved. Basic science is sometimes referred to as basic research.

1.2.2 Applied Research or Technology Development

Basic science is the foundation to applied research or technology development. As stated in its name, applied research begins to refine the application of the basic science (NSF, 1953). Applied research's goal is to take the underlying science discoveries from basic research and shape them into applications. While not fully into developing products, applied research bridges the gap from science to solutions.

1.2.3 Product Development

Research and development concludes with product development. This phase is the set of activities that integrates components, processes, materials, and applications into a product. The conclusion of product development is the product itself (Privitera, Design, & Johnson, 2009).

1.2.4 Technology Readiness Levels

As a technology progresses to fielding, it matures. This evolution is referred to as technology maturation. To better qualify the steps of technology maturation, the public sector has developed technology readiness levels. In the Department of Defense, Department of Energy, and National Aeronautics and Space Administration, a 9-tiered system of levels is used to describe the maturity of technology (GAO, 2016). These levels – Technology Readiness Levels – are used to assess the progress of technology in development (Azizian, Mazzuchi & Sarkani, 2011). They provide research and development management a standard template and consistent measurement for projects across their portfolio.

Some important levels of technology readiness include (a complete description of all technology readiness levels is in Appendix A):

- TRL 1 – This signifies the inception of the technology's progress. At this point, the technology is simply on paper, with a discussion on the observation basic underpinning scientific research. The risk at TRL 1 is qualified as "very high" ((Moorhouse, 2002). This marks the beginning of the transition from basic research to applied research.
- TRL 3 - At TRL 3, the technology achieves proof-of-concept, but does so at the component level without integration in a system (GAO, 2016). Modeling at this level is predictive, but not correlated with any operational usage or system intended configuration (Moorhouse, 2002). The risk at TRL 3 is qualified as "high" (Moorhouse, 2002).

- TRL 4 – At TRL 4, the technology reaches a level of operation. At this level, the technology must be demonstrated to operate in a laboratory environment although it is not in a form that resembles its final product. It may be an ad hoc connected network of components (GAO, 2016). Reasonable design and performance predictions may be possible with models from TRL 4 (Moorhouse, 2002). The risk at TRL 4 is qualified as "medium" (Moorhouse, 2002).
- TRL 6 – At TRL 6, the technology should demonstrate its function in a representative environment. This level is the first time that the technology operates in an environment that reflects how the product will be used after it is fielded. A prototype is constructed at this TRL (GAO, 2016). Organizations request TRL 6 or greater prior to inclusion in a formal acquisition program. Defense Department systems that had TRL 6 or higher maturity technologies prior to entrance into acquisition performed better with respect to cost and schedule (GAO, 1999).
- TRL 9 – At TRL 9, the technology reaches the conclusion of its evolution. It is then tested in an operational environment in its operational configuration. Operational testing and evaluation occurs at TRL 9 (GAO, 2016).

1.3 Focus

Intersecting the research and development categories with the technology readiness levels can help characterize the technology maturity segment that will be addressed. As seen in Figure 1-1 below (from D'Amico, O'Brien and Larkin, 2013), the "proof-of-concept" efforts flow from TRL 1 through TRL 4. These are aligned with the basic research and applied research phases. As described above, the performance parameters in this early stage is founded in modeling or none at all – showing a difficulty to quantify the performance (Moorhouse, 2002). The hallmarks of both phases, and these TRLs, is a higher degree of risk and the need for knowledge acquisition to mature the technology (Eckhause et al., 2009; GAO, 1999). "In essence, knowledge supplants risk over time." (GAO, 2012).

The United States invests more in research and development than any other country in the world. In 2013, the federal and state governments spent $133.5 billion on research and development. Importantly, 60% of that was spent on basic scientific research (Cannon, Ulferts, & Howard, n.d.). As stated by Kahn and McGourty (2009), "S[cience] & T[echnology] is difficult to measure in a meaningful way." Correlating with TRLs, science and technology can be viewed as TRLs up to 5/6 ((Moorhouse, 2002). With so much being spent on basic scientific research, the public sector needs tailored methodology to monitor the technical performance of those efforts. This will be the focus of the following research.

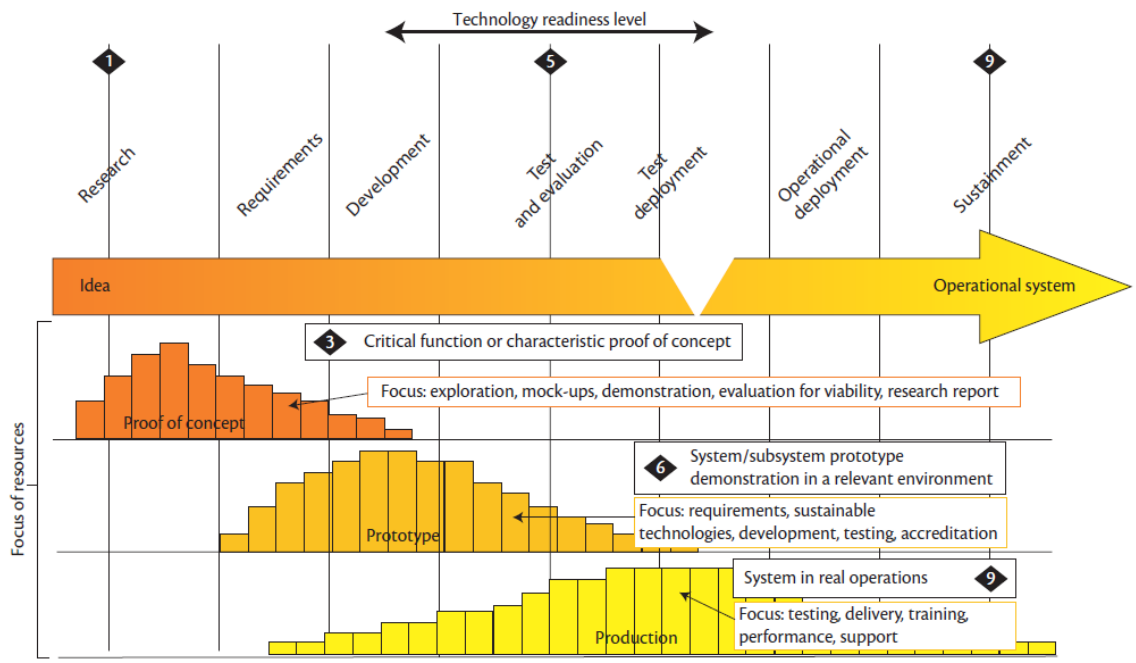

Figure 1-1: Technology Readiness Level and Focus of Resources chart

1.4 Problem Statement

Early stage research and development has technical performance parameters that are difficult to quantify. Due to this, current project management and systems engineering methods may not be optimized to measure technical status for research and development projects. This paper will recommend a conceptual framework for adapting Technical Performance Measurement (TPM) methodology for early stage research and development projects.

1.5 Scope of Study and Research Goals

The scope of this study is to map out a conceptual framework for implementing TPM for early stage research and development. It will synthesize concepts from Risk Value Method, Project Management, Analytical Hierarchy Process, and the statistical tool of segmented linear regression into Learning Performance Measurement. The research will test the predictive capability of the conceptual framework.

This research will review the currently accepted management processes and identify the gaps in their capabilities for early stage research and development. It will show how the concept of tracking progress by risk degradation, used in the Risk Value Method, can be adapted for tracking technical uncertainty. Using the project schedule and pairwise comparisons, it will devise a learning baseline for a real world project. Using segmented line regression, the research will predict future learning values for a simulated project. Finally, the research will test the accuracy of those predictions using a 95% predictive interval.

The culmination of the research will add another tool to the system engineer's or project manager's toolboxes. The broader research and development community will be provided a tool to assess and report progress outside of the normal boundaries of EVMS. They, collectively, will have available a way to explain progress on their project goal, outside of technical explanations or qualitative terminology.

1.6 Organization of Paper

This praxis began by introducing the problem and bounding its focus. It continued the introduction by proposing a problem statement and defining the scope of study. The introduction concluded with an explanation of the contribution of the research.

During the literature review, concepts will be explained to provide the foundation for the proposed methodology, as well as give a current state of research on topics affecting the problem. Applying those concepts, the research will propose a methodology to solve the problem, and then explain how that methodology was used in an experiment to test its utility.

Results of that experiment, analysis, and conclusions complete the research.

2 Literature Review

2.1 Systems Engineering

Systems Engineering is a natural discipline to address management of research and development. It is a discipline concerned with all aspects of a system over that system's life (INCOSE, 2017). According to the International Council on Systems Engineering (2017), the definition of systems engineering is:

> Systems Engineering is an engineering discipline whose responsibility is creating and executing an interdisciplinary process to ensure that the customer and stakeholder's needs are satisfied in a high-quality, trustworthy, cost-efficient and schedule compliant manner throughout a system's entire life cycle. This process is usually comprised of the following seven tasks: State the problem, Investigate alternatives, Model the system, Integrate, Launch the system, Assess performance, and Re-evaluate. These functions can be summarized with the acronym SIMILAR: State, Investigate, Model, Integrate, Launch, Assess and Re-evaluate. It is important to note that the Systems Engineering Process is not sequential. The functions are performed in a parallel and iterative manner.

As an interdisciplinary process, systems engineering takes advantage of other management processes and tools, such as those used in Project Management (Sharon, de Weck, & Dori, 2011). Additionally, systems engineers have developed tools for managing research and development. A component of those tools is Technical Performance Measurement (TPM), for management of the technical achievement of research and development (Sears & Taylor, 1984).

2.1.1 Technical Performance Measurement

TPM was developed by the Department of Defense to give managers a method to track and analyze technical performance, including manage predicted future values and analyzing variance from baseline (Sears & Taylor, 1984). The Technical Performance Measurements Handbook (Sears & Taylor, 1984) advocates that this management process should be structured by four key elements:

- “Status against a plan
- Estimate of future attainment
- Variance analysis
- Problem identification analysis”

TPM utilizes measurements of specific parameters to ensure the system or research and development project is progressing adequately with respect to its technical performance (Sears & Taylor, 1984).

2.1.1.1 Developing Measurements

In order to accomplish the tracking of the technical performance, TPM relies on the synthesis of various measurements to describe the requirements of the system[1] in “Technical Performance Measures” (Roedler & Jones, 2005). Individual TPMs are most directly derived from the system’s measures of performance (MOP) (Roedler & Jones, 2005). MOPs are, as named, measures for performance of the system, expressed numerically. They are the functional factors of the physical aspects of the system (Roedler & Jones, 2005). They are themselves derived from Measures of Suitability (MOS), and Measures of Effectiveness (MOE), as well as Key Performance Parameters (Roedler & Jones, 2005).

MOSs describe the desired states for the system’s maintainability, environments, operational factors, readiness and other requirements that affect the system’s operation, not necessarily performance (Roedler & Jones, 2005). MOEs describe the requirements for the system’s mission accomplishment (Roedler & Jones, 2005). KPPs establish threshold values for mission accomplishment, performance, and operational factors for

[1] “System” is used interchangeably with “research and development project” or “project” to describe an effort to develop a system.

the system, and can be a synthesis of the MOEs and MOSs (Roedler & Jones, 2005). They have a great impact on the system's overall effectiveness and suitability, if not met (Roedler & Jones, 2005).

The result of all these is a cascading hierarchy of measures that are distilled into the Technical Performance Measures (Roedler & Jones, 2005). These are the measures that are tracked along the system's maturity until the end of its development (Roedler & Jones, 2005). They include suitability and effectiveness measures, so TPMs include operational capability and quality factors (Roedler & Jones, 2005). For example, the TPMs for an Unmanned Aerial Vehicle (UAV) may be range, payload, and periodic maintenance requirements.

While they describe different aspects of the system, the TPMs may also be related (DAU, 2017b). Referring to the UAV example, increased payload may decrease range and vice versa. These two TPMs describe individual aspects of performance, and are related.

2.1.1.2 Implementing TPM

Once the TPMs have been decided, and the research and development activities have begun, TPMs' utility to management is tracking those measurements toward their goals. In this, the TPMs are assessed at regular intervals and compared to a baseline (Roedler & Jones, 2005). That baseline is established during the planning phases of the system (Sears & Taylor, 1984). When a variance in performance from that baseline occurs, assessments are done to analyze the impact and reasons for those variances (DAU, 2017b).

To bound the decision making of the program managers and systems engineers, an upper and lower limit for the TPM are established (Roedler & Jones, 2005). The “tolerance band” formed by these two limits serve as expectations for the “achieved-to-date” and “predicted value” values for the managers (Roedler & Jones, 2005). All these values push toward a “maximum requirement threshold,” or the desired requirement for that TPM (Roedler & Jones, 2005). All these values are plotted on a graph of time versus the measurement value (Roedler & Jones, 2005). An example of this tracking is shown in Figure 2-1, taken from the Technical Performance Measurement webpage by the Defense Acquisition University.

As illustrated by the small solid triangles on the X-axis in Figure 2-1 (from DAU, 2017b), one of the key features of the DOD’s acquisition system, and utilized in systems engineering and TPM, are technical reviews or audit. These may also be called milestones (DAU, 2017a). These technical reviews are not driven by schedule, but are accomplished when a certain achievement or event occurs (DAU, 2017a). The technical reviews, as systems engineering processes, are focused on technical aspects of the system, degradation of risk, technological maturity, and other systems aspects compared to the baseline plan (DAU, 2017a).

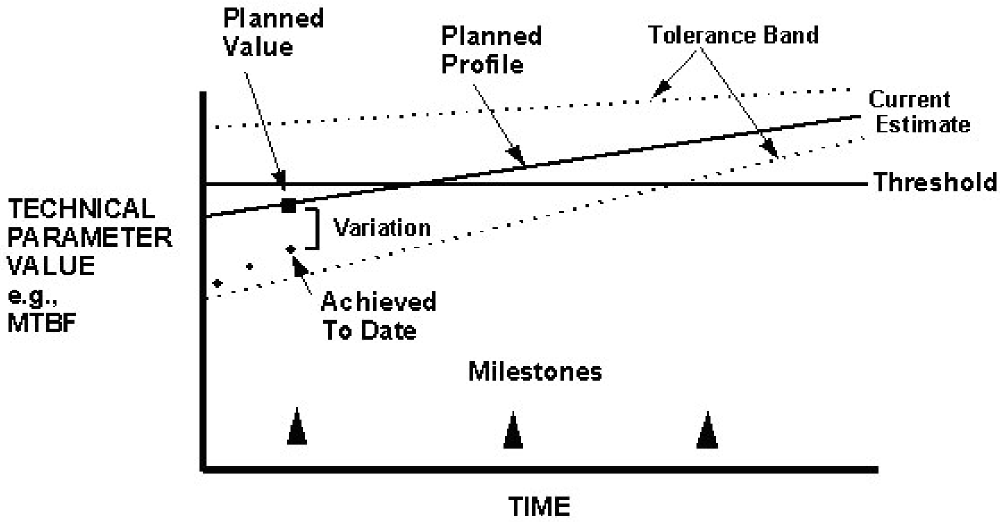

Figure 2-1: Example TPM Tracking Chart

Technical reviews are tailored specifically to meet the needs of the system (DAU, 2017a). Various government agencies direct execution of specific reviews at specific gates, but the non-mandatory reviews are made for that individual system (DAU, 2017a). They are executed when requisite levels of knowledge are achieved and are documented in the Systems Engineering Plan (DAU, 2017a).

Technical reviews also have an important role in future planning. They are intended to review the progress of the system in the context of what the progress was planned to be (DAU, 2017a). Exiting the technical review, a new baseline is established for the system, as it is the iterative product of assessing the technical progress, risk, schedule and taking a predictive approach to how the system will react in the future (DAU, 2017a). New dates for technical reviews can be established at this time (DAU, 2017a).

Technical reviews are tailored for the system (DAU, 2017a). Because of that, they can be executed on the system or any of its sub-systems, or a physical aspect of the system (DAU, 2017a). Individual measures or aspects of the system can be focused on to review during one of these incremental reviews (DAU, 2017a). This can also allow for increased scrutiny on interdependencies, or simply increased scrutiny on that aspect of the system (DAU, 2017a). Should these occur, they would likely lead the complete system technical reviews (DAU, 2017a).

2.1.2 TPM Alternative – TRI

Various implementations of TPM exist that integrate or use risk as a measurement tool (Garvey & Cho, 2005; Mahafza, Componation & Tippett, 2005). These methods vary in how they calculate, manipulate, or consider risk. They are above and beyond normal TPM, which is a natural basis for discussion of risk, as they expand on that interaction and capture risk numerically.

TPM Risk Index is a method that describes "how individual TPMs may be combined to measure and monitor the overall performance risk of a system (or SoS)" (Garvey & Cho, 2005). The TRI methodology advocates a method to measure TPM risk, normalize it into a dimensionless value, and then aggregate it into the overall system risk (Garvey & Cho, 2005). By doing this, TRI can help highlight which TPMs are leading the system risk, how risk changes over time, and overall provide management a method to assess and mitigate risk (Garvey & Cho, 2005).

TRI divides TPMs into two categories, *A* and *B*. These are simply defined as TPMs that decrease in value to reach their threshold (Category A), and TPMs that increase in value to reach their threshold (Category B) (Garvey & Cho, 2005). This is an important distinction, as it is required to normalize the risk values.

TRI begins its quantification by normalizing the TPM thresholds and measurements. Basically, the TPM measurements are subtracted by the threshold value for that TPM and either reduced or increased by 1 for Category A and B, respectively (Garvey & Cho 2005). That value is then divided by the threshold value to achieve the normalized, dimensionless measurement for TRI (Garvey & Cho 2005).

Once the TPMs have been normalized, the risk index is calculated. Weighting the individual risk values that were previously calculated allows the individual TPMs to be combined for the system value (Garvey & Cho 2005). The weighted risk values are all added by category, and provide the upper and lower bound of the TPM Risk Index (Garvey & Cho 2005). Once this is calculated, the management has a baseline and boundaries to assess system risk (Garvey & Cho 2005).

2.1.3 TPM Risk Index Alternative - TPRI

Various alternative implementations to TRI also exist. This includes the TPM Performance Risk Index. TRI calculates deviations from baseline and normalizes those to assess system risk. However, it does not necessarily calculate, or measure risk, as it derives its values from performance (Mahafza et al., 2005).

To alter TRI to add a measurement of risk, Mahafza, Componation, and Tippett (2005) developed a TPM Performance Risk Index that introduces a Degree of Difficulty (DD) modification. DD is a qualitative scale of degrees of difficulty that are represented by 0 to 6 levels. Level 0 is a 0 value, Level 1 is 0.1, and each level increases by 0.2 thereafter. The levels range from “No Risk” to “Guaranteed Failure” (Mahafza et al., 2005).

2.1.4 TPM Alternative – Risk Value Method

Browning, Deyst, Eppinger, and Whitney (2002) introduced a methodology for product development (PD) that essentially measures the degradation of risk as a gauge of adding customer value. This methodology is based on the architecture of TPM and integrates risk assessment to make its measurements. The Risk Value Method uses a measurement of the technical performance risk integrated with TPM to show degradation of risk, adding customer value, and therefore, progress on the PD (Browning, Deyst, Eppinger, & Whitney, 2002).

Browning, Deyst, Eppinger, and Whitney (2002) define technical performance risk as “an/the uncertainty that a product design will satisfy technical requirements and the consequences thereof.” Technical performance risk is the essential component to this method. The Risk Value Method requires an assessment of uncertainty and consequences to calculate the risk for each TPM (the technical performance risk), and then combine them to arrive at the overall project performance risk (Browning et al., 2002).

Risk Value Method uses a probability density function (PDF) to both quantify and normalize the value of uncertainty (Browning et al., 2002). It uses a PDF to represent the relative probability of possible outcomes of the individual PDFs (Browning et al., 2002). The shape of the PDF is dependent on the relative probabilities of the outcomes and where those outcomes exist in relation to the requirements (Browning et al., 2002). Acknowledging that making relative probabilities may be difficult, the Risk Value Method uses a triangular PDF (TriPDF) when information about the estimates is low (Browning et al., 2002). This TriPDF is a triangular PDF that is made by the worst case, most likely, and best-case estimates (Browning et al., 2002). The area underneath the TriPDF or PDF is normalized to 1 (Browning et al., 2002).

Calculating the consequences of each TPM is a process between the customer and the management team (Browning et al., 2002). The consequences are not necessarily represented by a simple linear function or quadratic function that describes the relationship. To represent the relationship of changes in the TPM value and impact to the customer, the Risk Value Method uses a utility curve (Browning et al., 2002). A utility curve can more accurately describe the customer value. For example, it may follow values along the unacceptable range, spiking quickly at a threshold value, then growing past the threshold while increasing as shown in the Figure 2-2 (from Browning et al., 2002) below. Figure 2-2 shows how an increase in the UAV range (in nautical miles) on the X-axis affects the normalized value of the customer value and consequence on the Y-axis.

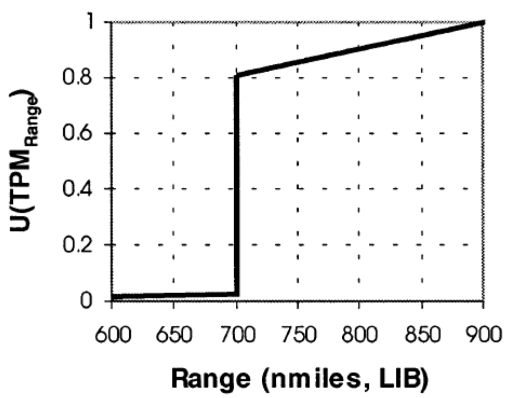

Figure 2-2: Example utility curve

After the utility curve is made, the consequence is determined by the difference between the target and the outcome of the TPM. In order to make this value meaningful, it is modified by a constant that changes it to a normalized value (Browning et al., 2002). The examples cited by Browning et al. are units to be purchased, profit in dollars, etc. (2002). The consequence is represented the following equation:

$$I_{\mathrm{TPM}} = \kappa_{\mathrm{TPM}} \left[U_{\mathrm{TPM}} \left(T_{\mathrm{TPM}} \right) - U_{\mathrm{TPM}} \left(x_0 \right) \right] \tag{2-1}$$

Where $\mathscr{K}$ is the constant, U_{TPM} is the utility curve, T_{TPM} is the target, and x_0 is the outcome (Browning et al., 2002). The practical interpretation of Equation 2-1 is that the difference in the target and the outcome times the constant provides the normalized consequence. The target and outcome are plotted on the utility curve, so their difference reacts according to that relationship.

Equation 2-2, represents the performance risk for a TPM. $\mathscr{R}_{TPM}$ is the performance risk for a TPM, and f_{TPM} is the PDF of all the outcomes of the TPM (Browning et al., 2002). The latter portion of the equation is the consequence from Equation 2-1. Equation 2-2 uses the probability of occurrence, through the integral, with the consequence times the constant to develop performance risk for each TPM. When there are a discrete number of outcomes, the integral in the equation is approximated (Browning et al., 2002).

$$\mathscr{R}_{\mathrm{TPM}} = \kappa_{\mathrm{TPM}} \int_{-\infty}^{T_{\mathrm{TPM}}} f_{\mathrm{TPM}}(x_0) \cdot \left[U_{\mathrm{TPM}}(T_{\mathrm{TPM}}) - U_{\mathrm{TPM}}(x_0)\right] dx_0 \tag{2-2}$$

Determining the overall product performance risk is an aggregation of the risk equation. As used by other methods, the individual TPM risks are weighted to make their values impact the overall product performance risk in accordance with their importance. With any weighted set, Risk Value acknowledges that large weighted or large risk/consequence $\mathscr{R}_{TPM}$ could overwhelm the other $\mathscr{R}_{TPMs}$ (Browning et al., 2002). This situation underscores the importance of viewing the risks holistically if possible (Browning et al., 2002). The overall product performance risk, $\mathscr{R}$, is the sum of the weighted ($\boldsymbol{w}_i$) individual TPM risks and is represented by the following equation:

$$\mathscr{R} = \sum_i w_i \, \mathscr{R}_{\text{TPM}, i} \tag{2-3}$$

Browning, Deyst, Eppinger, and Whitney (2002) also discuss linking activities to their impact to TPMs. They define activities as “broadly any effort resulting in new information, including decisions and reviews” (Browning et al., 2002). Risk Value advocates developing an *Activity-to-TPM* table that maps each activity to its impact on the TPMs (Browning et al., 2002). These tables serve as tools to review workload across the TPMs, see which activities do not contribute or have impact on TPMs, and can inform cost and schedule decisions (Browning et al., 2002).

2.1.5 TPM shortfalls for early stage research and development

TPM may not be best suited for measuring progress toward proof of concept, TRL 4, or other situations in technology maturation where the technical measures are difficult to quantify. At the early stages of the technology, the technical aspects of the system may not be readily measurable as they are yet to be proven (GAO, 2016). Up to TRL 4, even modeling and simulation may not reflect performance (Moorhouse, 2002). The measures for quality are unnecessary and too early in the project to be needed, as the technology is yet to be proven. Given that the TPM alternatives including TRI, Risk Value Method, and TPRI use the same measurements of technical performance, they too are susceptible to this shortfall.

2.2 Project Management

Projects by definition are “temporary endeavor(s) undertaken to create a unique product, service, or result” (Project Management Institute [PMI], 2013). Organizations have adopted project management tools, processes, and products to manage their activities that meet this definition. This project management culture can infiltrate all aspects of management – including managing innovation. (Meredith & Mantel, 2015, p.8)

2.2.1 Project Planning

In project planning, the blueprint for project execution is made (PMI, 2013). The work breakdown structure (WBS) is completed to define deliverables and activities needed to complete those deliverables (Cleland, 2004, p.104). The WBS is a source document for identifying TPM precursors, which makes it an important step in TPM (DAU, 201b). How long the project will take, its activities, and the resources to achieve success are documented (Cleland, 2004, p.104). Various component plans are completed including communication plan, risk management plan, etc. All these documents serve as the foundation for the execution phase where they will be implemented (Meredith & Mantel, 2015).

2.2.2 Monitoring and Control

The primary function of monitoring and control is to understand the project status, in order to apply corrective action when that status deviates from the project plan (Meredith & Mantel, 2015, p.435). Monitoring is the function of assessing project status, whereas control is the application of resources (PMI, 2013.). This occurs throughout project execution. Various systems have been devised to monitor project status, to be discussed in this section.

Monitoring project status is an imperative part of project management. As dictated in project plans, they may be calculated at regular intervals such as weekly, monthly, or quarterly (Meredith & Mantel, 2015, p.462-463). The capability for reporting has been increased through the usage of project management software, such as Microsoft

Project © which can automate some of the calculations (Meredith & Mantel, 2015, p.462-463).

2.2.2.1 Earned Value Management System

The Earned Value Management System was first popularized in 1962 in the Department of Defense (Kerzner & Saladis, 2009). Since that time, EVMS has become a recommended technique for monitoring project status, as described in the Project Management Body of Knowledge (PMI, 2013). Fundamentally, EVMS distills work, schedule, and costs into metrics to analyze current and future project performance (Kerzner & Saladis, 2009). Critical to the execution of EVMS is how the value is quantified. According to Khamooshi and Golafshani (2014), “a clear project scope, a well-defined schedule, and a detailed budget lay the foundation for implementing EVM and its derivatives for a project.”

In some EVMS applications, the planned value and earned valued are calculated by dividing spent time and cost (Meredith & Mantel, 2015). The spent time is also subject to different methods for calculation. These “earning rules” can be calculated in various ways, including the “50-50 rule,” which credits a task on the schedule 50% of the days at the start of the task and the other 50% at completion (Meredith & Mantel, 2015, p.451). The “0-100 rule” and “20-80 rule” are executed similarly, albeit with different percentages. The “critical input rule” can describe a task progress by the usage of a critical input (Meredith & Mantel, 2015, p.451). For example, task credit would be earned by the number of square feet of drywall used in the internal building of a house. The proportion earned would be determined prior to project start. Further rules are available at the determination of project management (Meredith & Mantel, 2015 p.451).

As important as it is to measure project status, EVMS is also an important starting point to forecast future values. Various methods exist to forecast cost and schedule based on previously accomplished data (Anbari, 2003; Batselier and Vanhoucke, 2015; Batselier and Vanhoucke, 2017; Hong, 2014).

2.2.2.2 Alternative EVMS Implementation – Level of Effort

One criticism of EVMS is that for projects with level of effort tasks, its cost performance measure can be inaccurate (Townsend, Mazzuchi and Sarkani, 2014). Level of Effort (LOE) tasks are those that do not end in a deliverable or lack discrete measurements (Townsend et al., 2014). Engineering (which include some research and development) projects can have numerous level of effort tasks, which would make EVMS an inaccurate tool for their measurement (Townsend et al., 2014). With these tasks in a project, the EVMS schedule and cost performance give a constant result, which is unlikely to lead to an accurate status. To mitigate this shortcoming, some approaches have been developed (Townsend et al., 2014).

One method involves separating LOE tasks from Discrete Effort tasks, those that end in a measurable deliverable, so that LOE tasks don’t influence the EVMS status (Townsend et al., 2014). The rationale being that the deliverables, DE tasks, are the most important. A further method renames LOE to Operational Effort (OE), and then separating from DE at the task level (Townsend et al., 2014). The tasks are then measured on separate vectors. At the project level, OE can overcome the DE portions and the project data becomes inaccurate (Townsend et al., 2014).

A less direct method to address LOE in project management monitoring involves using the project baseline to calculate bounded values (Townsend et al., 2014). By removing the remaining work hours from the baseline and bounding with 0, the method derives a relationship for LOE (Townsend et al., 2014). The result is a bounded value the project manager can use to compare the project performance. This method, however, fails when the projects are shorter and is only seemingly effective when the value approaches the baseline (Townsend et al., 2014).

A fourth recommendation to monitor projects with LOE tasks is using a similar indirect methodology as the project baseline method (Townsend et al., 2014). This method refers to planned value (PV) and earned value (EV) as equal when the task staffing is sufficient (Townsend et al., 2014). When the LOE task is insufficiently staffed, the EV drops to less than 1but when staffing is increased, the EV grows, but does not exceed 1 (Townsend et al., 2014). Using this relationship, the schedule performance measure will not be above 1. With that value for the schedule performance, the task would always show as favorably progressing (Townsend et al., 2014).

The fifth and final method for discussion is the one proposed by Townsend et al. in 2010. This method assumes that for engineering projects, there will be a database of action items for level of effort tasks (Townsend et al., 2014). By tracking the difference of the open and closed action items, and multiplying with the earned value, the earned value for level of effort is calculated (Townsend et al., 2014). Additionally, Townsend et al. proposes using this value to calculate a schedule performance index (2014).

2.2.2.3 Earned Schedule

Earned Value seems to be an effective method for monitoring the cost aspect of projects, but has been shown to be less effective for monitoring schedule (Lipke & Henderson, 2006). In order to monitor schedule, Earned Schedule method was developed.

In this method, EV and PV are mathematically manipulated to derive the schedule performance of the project (Lipke & Henderson, 2006). In this calculation, the cost factor is stripped and the result is schedule status. Practically speaking, the output is expressed in plain language (Lipke & Henderson, 2006). Earned Schedule also recommends various calculations to calculate future performance (Lipke & Henderson, 2006).

While shown to provide better project schedule status than EVMS, Earned Schedule relies on the backbone of EVMS for its calculations (Khamooshi & Golafshani, 2014). As such, the costs are considered then stripped, rather than focusing on schedule solely (Khamooshi & Golafshani, 2014).

2.2.2.4 Earned Duration Management

Earned Duration Management is an alternative monitoring method to EVMS. It can trace its lineage to EVMS, but prioritizes the movement of schedule as a measurement rather than value in terms of cost (Khamooshi & Golafshani, 2014). It assigns each task a value per unit of time (Khamooshi & Golafshani, 2014). For example, each task would be given a single unit for each day it is planned to be performed. If there are concurrent tasks, the values are added to get the total duration value for that day. If there are three tasks, each assigned a 1, the project planned duration would be 3 for that day.

2.2.3 EVMS Shortfalls for Research and Development projects

Research and development projects, particularly in the public sector, may be subjected to the EVMS requirements just as any other project. EVMS, and its derivatives, inherently lack a measurement of quality (Abdullah, Hamzah, Ismail & Razak, n.d.; Bower & Finegan, 2009). In this context, quality is analogous to technical performance.

Without a quality measure, it can be reasonably determined that EVMS is not effective for measuring technical performance. Frequently noted already, the high degree of uncertainty in R&D projects only expands the impact of that shortfall.

2.3 Risk Management

Risk management is a discipline interwoven in both project management and systems engineering (Sage & Rouse, 2009; Meredith & Mantel, 2015, Volkert, Stracener, & Yu, 2014). It is also an important consideration in innovation, as the risk due to unknowns tends to be high (Browning et al., 2002; Browning, 2014). Risk management includes many processes and tools including mitigation and transfer strategies, risk assessment, risk identification, and many others (Hubbard, 2009). The categorization of risk requires discussion to understand innovation risk (Williams, 1997).

There are many ways to divide risk into different categories, depending on the project, activity, or system. There does not appear to be a universal or standardized set of categories in project management or the broader systems engineering. For purposes of this problem, only a single category, technical performance risk, is addressed.

According to Browning et al. (2002), "Technical performance risk is the uncertainty that a product design will satisfy technical requirements and the consequences thereof." Savci and Kayis (2006) defined technical risk a bit more broadly and with more specificity:

> "Technical risk: related to a professional trade involving mechanical or industrial arts or the applied sciences. It includes design specific issues plus manufacturing specific issues such as quality assurance, product/process design, technological know-how, innovation, and technical support."

As determined by this definition, technical risk concerned with how the system or technology operates, or performs. It does not involve program issues such as budget, cost, or staffing. It does not concern project issues such as schedule.

2.4 Analytical Hierarchy Process

In the discipline of multi-criteria decision making, Analytical Hierarchy Process (AHP) was developed by Thomas Saaty as a tool (1986). AHP is usually used to explain priorities and is useful in down-selecting alternatives (Bible & Bivins, 2011). It relies on hierarchies of priorities that can be assessed using pairwise comparisons (Saaty, 1986).

In AHP, criteria are decomposed into different levels, so that they can be judged in smaller packages (Saaty, 1986). That decomposition results in a hierarchy that ends with the overall goal (Bible & Bivins, 2011). Each level of the hierarchy has a local weight of 1.00, but the individual elements are weighted according to their value when considered in the total hierarchy (Saaty, 1986).

AHP uses pairwise comparisons to both calculate the weights and assess alternatives against those criteria (Saaty, 1986). Once the hierarchy of criteria is calculated, decision alternatives are assessed against those criteria (Saaty, 1986). By finding how the alternatives score according to each criterion and aggregating them, a list of prioritized alternatives is created (Saaty, 1986). This list can be analyzed by many different lenses – isolating scores for individual criteria, etc.

2.4.1 Pairwise comparisons

The backbone of AHP is the execution of pairwise comparisons (Bible & Bivins, 2011). These assessments are used to analyze the relationship between individual criteria or alternatives resulting numerical weights for each (Bible & Bivins, 2011). The process can take subjective assessments and turn them into numerical scores (Saaty, 1986; Forman & Selly, 2001). AHP results in weights in a ratio scale (Forman & Selly, 2001). This allows weights to be calculated as a fraction of 100% (Harker & Vargas, 1987).

To start the pairwise comparisons, a questionnaire is designed in which each choice is paired against all others in consideration for that decision, with respect to the criteria (Bunruamkaew, 2012). To score the comparisons, the questionnaire should ask to score each choice against the others in subjective terms that correspond with numerical values (Saaty, 1986). Saaty (1986) recommended using language shown and scores similar to Table 2-1 below:

Table 2-1: Score to subjective preference table

Score	Preference
1	Equal Importance
3	Moderate Importance
5	Strong or Essential Importance
7	Very Strong Importance
9	Extreme Importance

Choice	**Extremely More**	**Very Strongly More**	**Strongly More**	**Moderately More**	**Even**	**Moderately More**	**Strongly More**	**Very Strongly More**	**Extremely More**	**Choice**
1										2
1										3
2										3

Figure 2-3: Sample questionnaire for pairwise comparisons

The questionnaire in Figure 2-3 shows an example of what a questionnaire could look like. There are alternative implementations with number choices, with guiding words to explain their meaning. In the figure, a mark on the left of event in the first row would indicate preference for Choice 1 over 2, while a mark on the right would be the inverse (Bunruamkaew, 2012).

From those choices, a matrix of values can be developed (Bunruamkaew, 2012). The matrix is made by translating the comparisons into numerical values, such as those in Table 2-1. The matrix is filled in going across the rows. If the comparison favors the choice in the row, the whole number corresponding to the preference is filled in for that column. If the choice on the column is favored, the inverse of the preference number is filled in (Bunruamkaew, 2012). Table 2-2 shows an example of this.

Table 2-2: Example completed pairwise matrix

	1	**2**	**3**
1	1.00	3.00	3.00
2	0.33	1.00	5.00
3	0.33	0.20	1.00

Each column in the matrix should be totaled to start the calculation of the choice weights (Bunruamkaew, 2012). Each value in the matrix is divided by its column total to normalize the weights (Bunruamkaew, 2012). Once the normalization is complete, the average of the row values is the weight for that choice (Bunruamkaew, 2012).

To validate that the weights are consistent, the consistency ratio is calculated (Bible & Bevins, 2011). This is done by calculating the consistency measures and using those to calculate the consistency index (Bunruamkaew, 2012). The consistency measures are the product of the pairwise matrix by the weights column, divided by the weight for that choice (Bunruamkaew, 2012). The sum of all the consistency measures minus the number of measures, divided by one less than the number of measures achieves the consistency index (Bunruamkaew, 2012).

The consistency index divided by the random index garners the consistency ratio (Bunruamkaew, 2012). The random index values were developed by Saaty and are readily available. The consistency ratio should be less than 0.1 for acceptable consistency, but ratios below 0.2 are tolerable (Pedrycz & Song, 2011). Ratios of 1.0 are considered random and the assessment should be redone (Bible & Bivins, 2011). Figure 2-4 displays an example of a completed matrix.

	1	2	3	Weights	Consistency Measure	
1	0.60	0.71	0.33	0.549	3.462	
2	0.20	0.24	0.56	0.331	5.107	
3	0.20	0.05	0.11	0.120	3.723	
Total	1.00	1.00	1.00	0.451	0.549	**CI**
					0.580	**RI**
					0.946	**C. Ratio**

Figure 2-4: Completed example weight matrix with consistency ratio

2.5 Regression

2.5.1 Linear Regression

Regression analysis are methods in statistics used to estimate the relationships in sets of data (Hair, 2006, p.163). Numerous different regression methods exist to deal with data sets. They include linear and non-linear methods (Hair, 2006, p.163). Linear regression estimates the data's relationship by drawing a line through the data that minimizes the distance of each of the data points (least squares method) (Hair, 2006, 163).

2.5.2 Segmented Linear Regression

Some non-linear data can be regressed by breaking it into segments and making linear regressions on those segments. This method, called piecewise linear regression or segmented linear regression creates individual linear regressions that explain how the data in those segments behave (Ritzema, 1994). The points which join the individual segments are called “breakpoints” (Ritzema, 1994).

The individual linear segments have all the characteristics of a linear regression including: r-squared, slope, intercept, standard deviation, standard error of the estimate, etc. (Hair, 2006, p.163). While commonly used to assess the “goodness” of the regression, the coefficient of determination or r-squared is only one value to consider (Albright, Winston, & Zappe, 2011). The standard error of the estimate is another value to consider, or the standard deviation of the residuals (Albright et al., 2011). The greater the standard error of the estimate, the larger the spread of the data (Albright et al., 2011).

2.5.3 Prediction and Confidence Intervals

The individual linear regressions from the segmented linear regression can be assigned confidence intervals, as in any linear regression. Confidence intervals are bounds, upper and lower, that represent the confidence that a given percentage of data points are expected to fall in those bounds given a sample (Leininger, 2013).

When predicting data, based on a regression, a prediction interval is determined. That prediction interval is similar to the confidence interval, but since the prediction is based on data points not in the existing sample of data points, it contains a factor of uncertainty (Spence & Stanley, 2016). This interval is wider than the confidence interval, due to that factor of uncertainty. Its width is dependent on the percentage of data expected to fall within it, i.e. a 95% prediction interval is wider than a 90% (Savelli, 2013; Leininger, 2013)

To calculate prediction intervals, equation 2-4 (Leininger, 2013) below is used. Should confidence intervals be needed, Equation 2-5 is to be used (Leininger, 2013).

These two equations are to be the same, but it should be noted that the prediction interval contains an additional of 1. This is the uncertainty factor that increases the width of the interval (Leininger, 2013). Again, the prediction interval bounds a predicted point not in the data sample, while the confidence interval bounds existing data points in the sample (Spence & Stanley, 2016).

$$\hat{y} \pm t^{\star}_{n-2}\, s_y \sqrt{1 + \frac{1}{n} + \frac{(x^{\star} - \bar{x})^2}{(n-1)s_x^2}} \tag{2-4}$$

$$\hat{y} \pm t^{\star}_{n-2}\, s_y \sqrt{\frac{1}{n} + \frac{(x^{\star} - \bar{x})^2}{(n-1)s_x^2}} \tag{2-5}$$

The terms in the equations are:

y = predicted value
t^{*}_{n-2}= t value with n-2 degrees of freedom
n = sample size
x^{*}= x value for the prediction
x bar = mean
S_y = standard error of the estimate
S_x = standard deviation of x

2.6 Forecasting

Forecasting future values is a commonly known goal for numerous industries including the financial sector and the evening weather. Project management research has been dedicated to forecasting future values for project duration, cost, and other factors (Hong, 2014). Forecasting for technical performance has included eliciting expert opinion. E-TRI relies on the technical performance in TPM (Eggstaff, Mazzuchi, & Sarkani, 2014). Project management methods rely on schedule variance, cost variance, earned duration, earned schedule, and other method-specific applications (Lipke, Zwikael, Henderson & Anbari, 2009; Batsailer & Vanhoucke, 2017; Batsailer & Vanhoucke, 2015).

Although those methods are reliant on specific metrics, basic forecasting for time-series methods have been developed (NIST, 2012). These statistical methods can be used to forecast values in the future, given past performance (NIST, 2012).

2.6.1 Linear Regression Forecast

Likely one of the simplest forecasting techniques is linear regression forecasting. This type of forecasting uses a linear equation from a linear regression to forecast values. Based on a previously calculated linear equation, future values are entered into the equation to predict future values (Dekking, Kraailkamp, Lopuhaa, & Messter, 2005). These forecasts are reliant on the data to continue behaving consistent with the linear equation (Dekking et al., 2005). These are called Autoregressive models (NIST, 2012).

2.6.2 Moving Average

To summarize past data, a smoothing technique called a moving average is used (NIST, 2012). This technique uses average of a certain number of data points in the time series to make a smoothed series (NIST, 2012). Additionally, the moving average can be weighted so that the more recent data points weight for more in the average (NIST, 2012).

2.6.3 Box-Jenkins (ARIMA)

Box-Jenkins method is a combination of the Autoregressive and Moving Average methods (NIST, 2012). It is also called the ARIMA method, as it stands for **A**uto**R**egressive **I**ntegrated **M**oving **A**verage (NIST, 2012). Box-Jenkins combines the linear regression with moving average to increase accuracy over either method individually (De Gooijer & Hyndman, 2006). In a basic Box-Jenkins implementation, there should be a large data set to base forecasts on (over 50 points) (NIST, 2012). The data should be stationary, with constant mean, variance, and lacking seasonality (De Gooijer & Hyndman, 2006).

2.6.4 Exponential Smoothing

Exponential smoothing is a type of weighted moving average that is used to forecast (NIST, 2012). For exponential smoothing, the relationship of the weights for the moving average is exponential (NIST, 2012). There are three types of exponential smoothing: single, double, and triple (NIST, 2012). Each of the types is suited to deal with specific types of time series (NIST, 2012).

The most basic exponential smoothing technique is single exponential smoothing. Single exponential smoothing uses an exponential weighted moving average to smooth the time series (NIST, 2012). The single exponential smoothing method uses a smoothing constant to weight the moving average (Albright et al., 2011).

Equation 2-6 below shows the basic exponential smoothing equation. In this equation, S_i represents a smoothed observation while the smoothing constant is α (NIST, 2012). The smoothing constant is between 0 and 1 (NIST, 2012). Representing the original observation is *y*. In this method, calculating a smoothed observation, S_t, is accomplished by "damping" the last original observation and adding the last smoothed observation, "damped" by the inverse of the smoothing constant (NIST, 2012). The smoothing constant can be calculated or left to a software solution to determine (NIST, 2012).

$$S_t = \alpha y_{t-1} + (1-\alpha)\, S_{t-1} \tag{2-6}$$

Double exponential smoothing is a method that is used on data that has a trend. By adding an additional constant in for consideration – γ – double exponential smoothing can factor the trend data (NIST, 2012). The two equations below, 2-7 and 2-8, show how γ is integrated into the double exponential smoothing calculation. Like α, the γ is between 0 and 1(NIST, 2012). In both these equations a new term is introduced, *b.* This value is calculated to start the smoothing, and various methods exist to calculate that value (NIST, 2012).

$$S_t = \alpha y_t + (1-\alpha)\,(S_{t-1} + b_{t-1}) \qquad (2\text{-}7)$$

$$b_t = \gamma\,(S_t - S_{t-1}) + (1-\gamma)\,b_{t-1} \qquad (2\text{-}8)$$

Finally, triple exponential smoothing, also called Holt-Winters after its developers, is more complex than both double and single exponential smoothing (NIST, 2012). While single exponential smoothing didn't consider a trend, double exponential smoothing did (NIST, 2012). Similarly, double exponential smoothing does not consider seasonality, so triple exponential seasonality does (De Gooijer & Hyndman, 2006).

Triple exponential smoothing uses three separate equations and additional terms (NIST, 2012). Due to this complexity, calculation of these terms is best left to a software solution (NIST, 2012). The below, Figure 2-5, is taken directly from the National Institute of Standards and Technology's Engineering Statistics Handbook under the heading "*6.4.3.5.Triple Exponential Smoothing*" (2012). As seen, the triple exponential smoothing equations are significant, which is why the recommendation is leave this calculation to a software solution.

$$S_t = \alpha \frac{y_t}{I_{t-L}} + (1-\alpha)(S_{t-1} + b_{t-1}) \quad \text{OVERALL SMOOTHING}$$

$$b_t = \gamma(S_t - S_{t-1}) + (1-\gamma)b_{t-1} \quad \text{TREND SMOOTHING}$$

$$I_t = \beta \frac{y_t}{S_t} + (1-\beta)I_{t-L} \quad \text{SEASONAL SMOOTHING}$$

$$F_{t+m} = (S_t + mb_t)I_{t-L+m} \quad \text{FORECAST ,}$$

where

- y is the observation
- S is the smoothed observation
- b is the trend factor
- I is the seasonal index
- F is the forecast at m periods ahead
- t is an index denoting a time period

Figure 2-5: Triple Exponential Smoothing equations and terms

2.6.5 Forecasting Comparison Methods

In order to compare differences in forecasting methods, various methodologies have been developed. The mean absolute deviation, or MAD, is used to compare the average of the absolute values of each forecast’s variation from the realized values (Hocking, 2013, p. 210). MAPE, or the mean absolute percentage error, performs a similar function to MAD, but does so in percentages (Montano Moreno, Palmer Pol, Abad, & Biasco, 2013). Doing so eliminates the scale and compares the percentage value (Montano Moreno et al., 2013).

The mean squared error or MSE, and root mean squared error, RMSE, both rely on the squares of the error (difference in actual and prediction) to access the accuracy. RMSE simply takes the root of MSE. In both methods, the smaller the value, the less error and therefore more accurate prediction (Thompson, 1990; Willmott & Matsuura, 2005).

3 Contribution

Given the investment the public sector commits to research and development, the public sector needs to manage early stage (TRL 4 and prior) research and development (Cannon et al., n.d.). It has the existing systems engineering and project management tools at its disposal to do so (PMI, 2013; INCOSE, 2017). The unique contribution to the body of knowledge proposed by this discussion is the modification of existing methodologies to develop a conceptual framework for assessing and predicting research and development project progress.

TPM, EVMS and their alternative implementations have been proposed and researched, as described in Section 2. Those methodologies have properties that make them suited to monitor technical and project progress, respectively. They also both have shortfalls with respect to management of early stage research and development. Primarily, EVMS is tailored to monitor project status, not technical progress (Abdullah et al., n.d.; Bower & Finegan, 2009). TPM is tailored to monitor technical status, when a defined level of technical achievement is already achieved (DAU, 2017b).

These two shortfalls for early stage research and development activities require consideration. In order to do so, the following research will describe how to address these issues. In the previous two sections, the problem was introduced and relevant literature was review. In the next sections, a proposal will be introduced. To guide discussion of that proposal, the following research question is introduced:

Research Question I

Can an uncertainty degradation method combined with existing technical performance measurement methodologies provide a framework for the tracking of progress and quantify expected future technical progress?

To answer this research question, a methodology will be developed and implemented on a real-world project. That new methodology, fully described in Section 4, is focused on monitoring the degradation of technical uncertainty for research and develop. The method is based on learning, using the concepts of TPM and will therefore be called "Learning Performance Management." This methodology was developed to use the concepts in Risk Value Method, but was tailored to measure uncertainty degradation rather than risk.

To guide the research toward answering the research question, three hypotheses were developed. First, a hypothesis that informed whether an uncertainty focused measurement could be developed for early stage research and development projects. This hypothesis guided the research to translate the idea to a real-world application. The first hypothesis is:

H1: An activity-based uncertainty measurement can be developed based on Risk Value Method concepts.

$H_0$1: An activity-based uncertainty measurement cannot be developed based on Risk Value Method concepts.

After demonstrating the ability to implement Learning Performance Management on a real-world project, the focus of the research will turn toward how the new methodology bounds the expectations of future performance. Having implemented the new methodology on a real-world project, with a defined baseline performance, the project schedule will be lengthened to simulate delays.

After simulating delays in the project, the Learning Performance Management's prediction process will be tested to assess if the realized measurements from the simulated delays fall within the calculated prediction intervals. This provides assurance to decision makers that the predictive capability of Learning Performance Measurement will behave within the normal uncertainty of 95%. This hypothesis reflects that test:

H2: Learning Performance Measurement predictions are valid within the 95% prediction interval.

$H_0$2: Learning Performance Measurement predictions are not valid within the 95% prediction interval.

Finally, the Learning Performance Measurement predictions were subjected to statistical scrutiny by hypothesis 2. Should those predictions be within the bounds of the prediction interval, they should be compared to other forecasting methods. Learning Performance Measurement predictions are unnecessary if no better than mathematical forecasting methods. To assess the effectiveness of the LPM forecasting against mathematical standard forecasting methods, the follow hypothesis will be tests:

H3: Learning Performance Measurement predictions perform better than mathematical standard forecasting methods.

$H_0$3: Learning Performance Measurement predictions perform worse than mathematical standard forecasting methods.

Having been subjected to three hypotheses, the research will be able to answer the research question. Should Learning Performance Measurement methodology pass the three hypotheses that are guiding the research, the research question can be affirmed.

4 Methodology

4.1 Overview and Approach

As described in Sections 2.1 and 2.2.2, many project management monitoring or performance management systems are designed to track work being accomplished toward deliverables. Earned Value Management System metrics use cost and completion methods to aggregate project progress (Meredith & Mantel, 2015). TPM tracks development projects progress toward particular aspects of requirements (Sears & Taylor, 1984). TPM Risk Index measures risk as of a system based on TPM measures (Gravey & Cho, 2005). The Risk Value Method monitors progress through the degradation of product performance risk. It does this by estimating relative probability (uncertainty) and consequence (Browning et al., 2002).

As previously discussed, the EVMS lacks a specified metric for quality (Abdullah et al., 2015). TPM, and therefore, TRI require that a technical baseline, with a defined starting technical performance value, be established to measure progress from (Sears & Taylor, 1984). The Risk Value Method requires defined consequences to derive its risk (Browning et al., 2002). All these factors may make it difficult to apply them to early stage research and development efforts, those prior to proof-of-concept, prior to validation in lab or relative environments, those research and development efforts whose consequences are difficult to define, or even those research and development efforts that do not have reasonable performance changes to measure.

By adapting the Risk Value Method, these difficulties can be overcome. A new implementation of the Risk Value Method concepts is proposed that focuses on uncertainty and indirectly manages consequence (Browning, 2002). Integrating project management tools and systems engineering tools, a method is proposed, Learning Performance Measurement (LPM), that will provide insight into technical progress when other research and development management systems are inadequate. At its most basic, LPM will measure the degradation of value-added technical uncertainty over time to provide insight into the technical progress of an research and development effort.

LPM uses the overarching concept of tracking progress via risk as advocated by Browning, et al. (2002). The Risk Value Method develops utility curves based on technical performance risk and customer value (Browning, 2002). As described in the definition of the TRLs, in early stage research and development, prior to proof-of-concept, the system characteristics are awaiting confirmation. Without confirmation, it can be assumed that specific customer value is yet to be determined.

For the following methodology, the customer value and the consequence are assumed to be binary values. Technical performance risk (uncertainty x consequence) can be seen as uncertainty measurement in early stage research and development, because the consequence is effectively 1. In basic science, consequence is completely lacking, because the application has not yet been determined. Prior to proof-of-concept, the consequence is whether the technology can be made to work or not.

In these situations, consequence is 0, concept not proved, or 1, proof-of-concept validated. The refinement of the technology occurs in latter TRLs, as shown in the TRL levels (GAO, 2016). If the consequence is 0 or 1, the customer value is also 0 or 1 for the same reasoning. Given these assumptions, the measurement of uncertainty degradation will provide a progress measure. However, how uncertainty is measured must be constrained in order to ensure it is value-added.

Browning, et al. (2002) and Browning (2014) would assert that degradation of uncertainty is not sufficient to measure progress because eliminating uncertainty that does not contibute to the research and development goal is not progress. Further, since research and development can be a highly non-linear, iterative process, uncertainty eliminated may be negated by new information (Browning, 2002). Browning, et al. (2002) uses consequence to overcome these issues. In order to address these issues, LPM puts certain constraints on the how it counts learning.

In the most basic sense, LPM only counts learning that degrades uncertainty toward the technical goal. Various activities may degrade schedule uncertainty, cost uncertainty, or other uncertainty. Those uncertainties are not counted as learning toward the technical goal. Activities may eliminate technical uncertainty by uncovering knowledge but it has no effect toward the technical goal. That uncertainty is not counted in LPM.

Additionally, activities may increase technical uncertainty about the goal rather than achieve it, due to the non-linear nature of innovation (Browning, 2002). That situation requires revisiting the planning assumptions and re-baselining the research and development efforts. That does not imply that those discoveries are ultimately waste. These discoveries may further contribute to the body of knowledge in other ways, but do require a rework of the schedule. These discoveries, new knowledge, are part of the overarching goals of research (OECD, 2015).

4.1.1 Learning Performance Measurement Implementation

Fundamentally, LPM is concerned with how much relative uncertainty the research and development effort is eliminating over the course of its execution. Just as the Risk Value Method uses the degradation of product performance risk, LPM uses the increase in “learning,” or the elimination of uncertainty toward the goal, as a measure of progress (Browning, 2002). Developing the learning baseline starts just as developing TPMs does, with deciding the necessary measurements (Sears & Taylor, 1984).

Deciding where learning needs to be tracked is not unlike deciding TPMs. Assessment of the WBS, project documents, and systems engineering documents allows for decisions to be made on where learning needs to be achieved. Effectively, the factors that would normally generate KPPs, MOSs, MOEs, and MOPs, are analyzed for where uncertainty needs to be eliminated, toward the goal. Individual TPMs are decided based on learning.

4.1.2 Developing Learning Values

As previously discussed, planning is an integral part of systems engineering, project management, TPM, and EVMS (PMI, 2013; DAU, 2017b). During planning, activities are planned to meet the goal of the R&D (PMI, 2013.; DAU 2017a). WBSs are developed that describe the resources, schedule, and activities needed to meet the goals of the project (Meredith & Mantel, 2015). As EVMS and TPM use these products to develop their measurements, LPM uses the activities to develop a learning baseline (Meredith & Mantel, 2015; DAU, 2017b).

In LPM implementation, the relative amount of uncertainty must be quantified. To develop this value, LPM will use pairwise comparisons and, depending on complexity, a hierarchy to make a relative learning baseline. Determining the exact amount of uncertainty is impossible, as R&D’s goal is to develop knowledge (OECD, 2015). It’s unlikely that an accurate value of the unknown could be determined.

Therefore, using the pairwise comparisons will develop a set of weights for each activity on a ratio scale (Forman & Selly, 2001). This provides the decision makers with uncertainty values proportional to the amount of uncertainty each is expected to satisfy.

The identification of learning can be similarly executed as making the MOPs for TPM, by decomposing the work breakdown structure to identify sources of learning (DAU, 2017b). Using a breakdown structure to identify risk is an effective method for bounding the problem, and quantifying risk (Fazli, 2013).

Similarly to Risk Value, the total system learning should be weighted to achieve a system value of 1 (Browning, 2002). This is an aggregate of how much technical uncertainty is expected to be eliminated toward the overarching goal of the system. Capturing learning will take a bottom up approach from the lowest WBS levels and add learning values, in a weighted fashion, up the hiearchy to achieve the system learning of 1.

Tackling a problem requires decompostion of the problem into simple elements. In that vein, and adhering to AHP principles, LPM uses the lowest levels to quantify technical uncertainty (Saaty, 1986). Starting at the bottom, LPM uses modified *Activity-to-TPM* tables from Browning's Risk Value Method to capture if activities contribute to learning and how that learning is achieved (2002). Deciding which activities contribute to learning is directly useful to make the learning values, while how the activities contribute to learning will be important in later steps (discussion to follow in Section 4.1.3). The activities used to develop the *Activity-to-LPM* are derived from the activities that populate a project schedule, or any planning artifact that describes how a system will mature its technology.

		Learning achieved...		
	Contribute to Learning	**Porportionally (even rate throughout task)**	**At the End**	**Other**
Activity 1	Y	X		
Activity 2	N			
Activity 3	N			
Activity 4	Y		X	
Activity 5	Y	X		
Activity 6	Y	X		
Activity 7	N			

Figure 4-1: Example contribution to learning table

As shown in Figure 4-1 above, the activities are listed for management, to mark those activities that contribute to learning and how they accomplish that learning. The list of activities that contribute to learning are then put into a questionaire or computer system to capture the pairwise comparisons.

Those activities that do not contribute to learning will not add value in this process, but retain necessity in the other project management and system engineering processes. The tasks that do not contribute to learning are those that are routine, well characterized, or simply required to progress the project goals. For example, one task in a project could be acquiring routine materials. This activity gains no knowledge, or learning toward maturing the technology, but is necessary to achieve the project goals. It is possible that many activities, or even WBS elements will not be considered for pairwise comparisons. An example of this is management tasks or WBS project management elements.

Figure 4-2 below shows an example of how the pairwise comparisons are completed. Numerous different methods and implementations of the pairwise comparisons can be imagined.

Amount expected to learn…										
Activity Name	**Extremely More**	**Very Strongly More**	**Strongly More**	**More**	**Even**	**More**	**Strongly More**	**Very Strongly More**	**Extremely More**	**Activity Name**
Activity 1			x							Activity 4
Activity 1								x		Activity 5
Activity 1						x				Activity 6
Activity 1					x					Activity 8
Activity 1		x								Activity 10
Activity 1						x				Activity 13

Figure 4-2: Sample LPM questionaire

Calculating the learning values for the system is accomplished using the AHP methodology in Section 2.4. By having that matrix, the process described in Section 2.4 can be used to calculate the weights of the activities.

Once the weights of the activities are calculated, the *Activity-to-LPM* (table 4-1) table is populated with those weights. An important note is that the example, Table 4-1 only shows a single measurement. It is possible that managers divide the table into many uncertainty measurements for pointed technical aspects, and that they may be displayed on this table, should they have the same activities. As described above, the *Activity-to-LPM* table displays the weights of learning and how the individual activities learn. If warranted, managers may choose to also display cost or schedule on this table to compare investment to return.

Table 4-1: Sample *Activity-to-LPM* table

Activity Name	Learning	How Learn
Activity 1	0.098	E
Activity 2	0.015	B
Activity 3	0.000	B
Activity 4	0.000	P
Activity 5	0.000	E
Activity 6	0.136	E
Activity 7	0.000	P
Activity 8	0.050	E
Activity 9	0.000	P
Activity 10	0.029	P
Activity 11	0.158	E
Activity 12	0.080	E
Activity 13	0.000	P
Activity 14	0.408	E
Activity 15	0.000	E
Activity 16	0.026	E
Activity 17	0.000	E
Total	**1.000**	

4.1.3 Developing the LPM Baseline

Using the activities to quantify the learning is effective for developing the learning weights, but also allows for developing a time-based learning baseline. Using the descriptor of how the activities learn, planners can assign the weights of learning to the activities, and use earning rules similar to those of EVMS to develop the expected learning baseline (Meredith & Mantel, 2015).

Essentially, how learning is achieved is tailored to whichever time scale may be needed for that individual activity's contribution to learning. For simplicity, the earning rules in project management tasks can be used as a template (Meredith & Mantel, 2015). Activities could learn proportionally to their schedule or time elapsed. An example may be a design task that provides new information as it progresses. Activities could learn at the end, or at the completion of the activity (Meredith & Mantel, 2015). Example of this type of learning could be a test that requires full completion before any data is acquired. Perhaps an activity is ten individual tests, with each contributing some learning. This may be a stepwise learning achievement. Many other situations could be envisioned with many other learning achievement scenarios (Meredith & Mantel, 2015).

Having defined how the activities will learn and the learning earning rules defined, the learning baseline is a simple project management task to map out the learning across the expected duration of the research and development effort. At each reporting period, the earned learning at that time is added to get the total learning. The total learning at each reporting period is charted for the life of the research and development effort. Figure 4-3 below displays an example of the learning baseline.

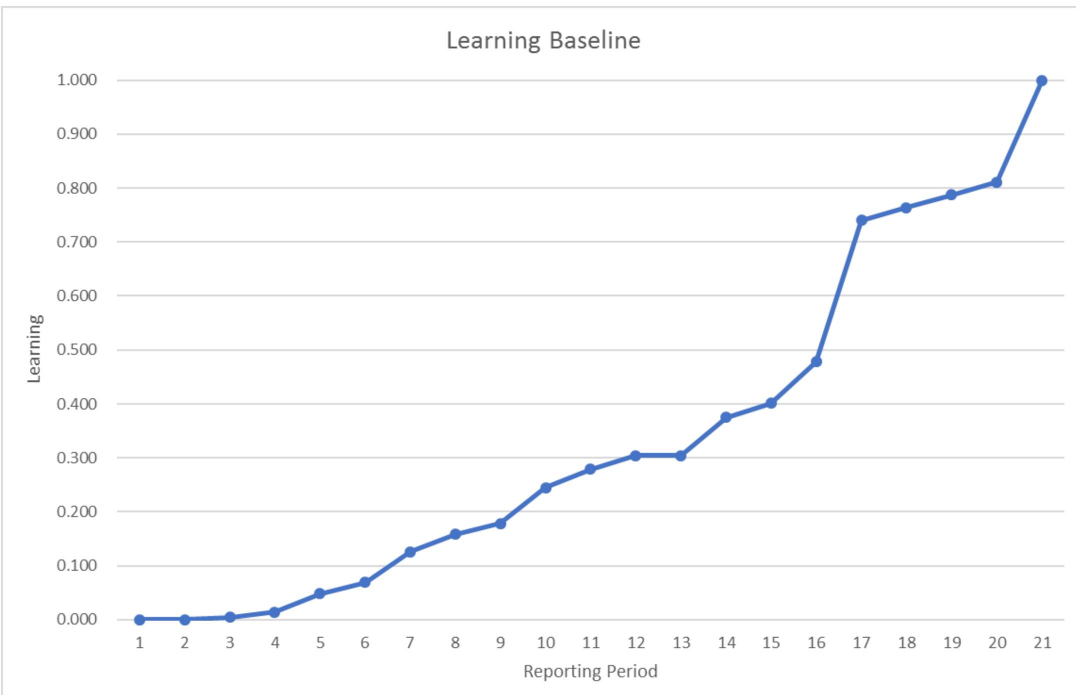

Figure 4-3: Example learning baseline

In execution, rewarding the amount of learning achieved requires the intervention of managers. The activities serve as the foundation for how much learning is credited for the research and development effort. Managers and technical staff are required to assess whether the activities contributed to learning as much as expected, less, or more. The expected conclusion of the research and development effort is that each activity is credited with the amount of learning in accordance with the baseline. It would also be possible that less learning is achieved and the R&D effort continues despite that information shortfall.

Since a defining characteristic of early stage research and development is a high degree of uncertainty, it is possible that the learning values are not consistent with the amount of uncertainty eliminated. An activity or period could uncover much more information than expected. Should the management team believe that the learning value in the baseline underestimates the amount of uncertainty encountered, a re-baselining can be executed.

Re-baselining the learning value is done in one of two ways. One way would be to simply execute the pairwise comparisons again. The other would be to add additional learning value to the activities that are uncovering more information than expected. This drives the learning value for the measurement over 1.0, and eliminates the utility of the ratio scale. However, it would explain the introduction of new uncertainty. Additionally, in this situation, the management team should be cognizant of the availability heuristic cognitive bias, which may cause them to favor recent data (Kynn, 2008).

4.1.4 LPM Technical Reviews

A universal necessity in systems engineering, project management, and the vast majority all management, is the need to periodically review the progress. In the context of a research and development effort, project, or system development when technology is maturing, there is the need to review technical performance. TPM supports these reviews, by providing the data necessary to review the progress (DAU, 2017a; DAU, 2017b).

LPM serves as the basis for executing the required technical reviews. The amount of learning achieved is used in lieu of a technical achievement or threshold value. Rather than choosing arbitrary intervals, a segmented linear regression is executed on the learning baseline. By doing this, the learning baseline is divided into linear segments that are joined at “breakpoints” (Ritzema, 1994). Those breakpoints are when the technical reviews are scheduled. This process divides the baseline into segments in which the research and development’ learning behaves similarly, i.e. similar rate of change from technical review to technical review (Ritzema, 1994).

Segmented linear regression can be easily performed by computational tools (Ritzema, 1994). However, these tools may not be accessible, or make breakpoints that do not meet the needs of the management team. The mathematically calculated breakpoints may be too infrequent or too near the beginning/end to be useful. For that reason, the learning baseline can also be manually regressed into segments with breakpoints that are adequately spaced, or meet other requirements of the management. This action is done by trial and error.

Manually dividing the baseline into linear segments is a simple regression with the learning baseline as the y-axis (dependent variable) and period as the x-axis (independent variable). The resulting regression gives segments with the linear equation with the y being learning and x being the period. This should be done starting at one breakpoint and ending at the next. To ensure a continuous set of values, each linear segment should include the previous breakpoint.

4.1.5 Predicting Future Values

The linear equation produced for each segment allows the prediction of future learning values by using the forecasting discussed in Section 2.6.1 (Dekking et al., 2005). In execution, a learning value is achieved for each period. The achieved learning value can be put in the linear equation for its segment to calculate at what time that value would have been achieved in the baseline. The expectation is that the values behave nominally consistent with the linear regressions calculated on the baseline, since the learning values are tied to the activities, which are tied to the project plan. This ties the baseline to the schedule logic. To start that process, the linear equation is solved for x, the period, using the value of learning achieved in execution.

By entering the achieved learning value in that equation, the output is the period for which the management team would expect to achieve that learning value, if the project were behaving in accordance with baseline, or the *expected period*. The *expected period* can then be used to calculate the *predicted learning* value, or learning expected to be achieved in the next period.

Given no intervention by management, (i.e. applying additional resources to return the schedule to its baseline), the project is expected to advance its learning – starting from the previously achieved learning value – at a rate constant within its segment, or consistent with the plan. The *expected period* indicates where the project would be at the learning achieved, so increasing the *expected period* to the next reporting period is used to calculate the *predicted learning*.

When learning values depart from each breakpoint, or technical review value, they should stop behaving consistently with the previous segment and pick up the behavior of the new segment.

4.1.6 Framing Decision Making

LPM predictions are simple mathematical calculations, based on planned values. They are not replacements for expert judgement or technical expertise. However, the predictions do provide an expectation of future accomplishments based on planning factors and observed performance.

Expectation of future achievement is uncertain. Fortunately, that uncertainty can be bounded using the prediction intervals described in Section 2.5.3. The width of the bounds of a prediction interval is an indicator of how uncertain the estimate is (Leininger, 2013). By integrating those bounds in the prediction of future learning, decision makers can see both how the research and development may learn in the next period, and how uncertain that estimate is (Savelli, 2013). Additionally, these prediction intervals can serve as the tolerance bands used in TPM, absent other established values (Sears & Taylor, 1984).

5 Experiment

5.1 Overview

In order to begin the assessment of LPM, gain insight during practical application, and answer the research questions, a project was selected to perform a trial implementation. Data was gathered from a public-sector technology maturation program. The program office developed a WBS for its entire portfolio. That WBS covered not only the technology maturation activities – development and qualification – but also manufacturing, modeling and simulation, program management, and implementation of the developed technology in five specific applications.

Since their portfolio covered many disparate projects and included implementation of developed technology, WBS was distilled to find an applicable project to assess LPM. One of the lowest level WBS elements was selected for the assessment. The project is reasonably simple, focused on a single technical goal, and the schedule and budget information was provided. The program office had previously adopted EVMS, but abandoned it as it did not add value for managing the project.

The project is a representation of how LPM could be implemented. The project aims to mature the technology and examine performance of a near-scale prototype in a laboratory environment. A TRL isn't defined, but this project would likely culminate in a technology readiness assessment of TRL 4. The project maturity is within the scope of the research. This project spans over nine and a half years, costs $6.4 million, and has not yet concluded.

TPM was not implemented, therefore, no technical performance measurements exist. Although, the project only produces results only at the near end and the modeling capabilities are yet to be validated as well, so TPM may not be particularly informative for assessing technical performance. As mentioned previously, management abandoned EVMS, as it was not decidedly effective in measuring performance, so EVMS data is informative but previously not found to be valuable for this project.

The research questions are answered by simulating the performance of the learning of the project. Ideally, the research would have been accomplished by implementing on a project, and watching the performance in real time to gather data. Unfortunately, due to the time constraints of the research, that was not possible. Given a year-long timeline, research, development of a methodology, and implementation were not possible. Simulating the performance of a project will provide a proof-of-concept analysis, founded in reality, and serve as starting point to additional research.

This simulation will be used to determine if there is an acceptable degree of uncertainty in predicting the future learning values, and developing insights in the LPM process. For the project, the learning performance baseline was calculated as described in Section 4.1.3 and the technical reviews were scheduled in accordance with Section 4.1.4. The project schedule was then altered by randomly lengthening the activity durations anywhere from 0 to 50%. New learning values were calculated for each reporting period based on that increased schedule.

Certain assumptions were made for the experiment:

- The simulated project status was unimpeded by management intervention. While delays and overruns would normally be calls for corrective action, they would introduce situational artificialities unimportant to the experiment (PMI, 2013). Specifically, planning assumptions are changed and the baseline is altered when management intervention occurs (PMI, 2013.).

- All activities eventually complete their scope. Again, should an activity fail, management intervention would be required.
- Tasks earn learning when complete or proportionally, as identified.

5.2 Establishing the Learning Values

For the project, the process in Section 4.1.2 was followed to determine which activities contributed to learning. Using the project schedule as a template, the figure below (Figure 5-1), was developed and sent to the technical program manager. The activities that contributed to learning were identified and how those activities learn. Certain "milestone" activities were omitted from this worksheet. These "milestone" activities had zero resources or duration, and serve as interim points in the project.

		Learning achieved...		
	Contribute to Learning	Proportionally (even rate throughout task)	At the End	Other
Conceptual Design	Y	x		
FSI Modeling Plan	N			
Conceptual Design Review (30%)	N			
Preliminary Design	Y	x		
60% Design Review	N			
Final Design	Y	x		
Final Design Review	N			
Irradiation Vehicle Fabrication	N			
Flow Testing	Y	x		
Prepare Characterization Plan	N			
Detailed Characterization	Y	x		
Characterization Summary Report	Y		x	
Safety Analysis	N			
Pre-Irradiation Material Properties	Y	x		
Integrated Test Assembly	N			
ATR-C Run	N			
Irradiation (Four ATR Cycles)	Y	x		
Cooling and Shipping	N			
As-Run Analysis	Y	x		
Irradiation Summary Report	Y		x	
Non-Destructive PIE	Y	x		
Destructive PIE	Y	x		
PIE Summary Report	Y		x	
Material Properties Summary Report	Y		x	
Post Irradiation Material Properties	Y	x		

Figure 5-1: Experiment Learning Contributions

During data gathering, the pairwise comparisons were completed in the same step as identifying the activities that contributed to learning. Those activities that didn't contribute to learning were simply not assessed. The pairwise comparisons were also completed by the technical program manager. The narrative of the comparisons was: "Extremely More," "Very Strongly More," "Strongly More," "More," and "Even." Appendix B shows the outcome of these assessments.

These comparisons were translated to numerical values and put through the process described in Section 2.4. These pairwise comparisons are calculated in the AHP process to get their relative weights to the rest of the activities (Bunruamkaew, 2012). These weights represent the relative amount of learning for each activity. The *Activity-to-LPM* table (Table 5-1) was completed to represent the learning values by activity.

Table 5-1: Completed *Activity-to-LPM* table for experiment project

Activity Name	**Learning**	**How Learns**
FSP-1 Conceptual Design	0.013	P
FSP-1 FSI Modeling Plan	0	
FSP-1 Conceptual Design Review (30%)	0	
FSP-1 Preliminary Design	0.012	P
FSP-1 60% Design Review	0	
FSP-1 Final Design	0.007	P
FSP-1 Final Design Review	0	
FSP-1 Design Complete	0	
FSP-1 Irradiation Vehicle Fabrication	0	
FSP-1 Flow Testing	0.014	P
FSP-1 Prepare Characterization Plan	0	
FSP-1 Detailed Characterization	0.033	P
FSP-1 Characterization Summary Report	0.038	E
FSP-1 Characterization Complete	0	
FSP-1 Safety Analysis	0	
FSP-1 Pre-Irradiation Material Properties	0.034	P
FSP-1 Integrated Test Assembly	0	
FSP-1 ATR-C Run	0	
FSP-1 Irradiation (Four ATR Cycles)	0.152	P
FSP-1 Cooling and Shipping	0	
FSP-1 As-Run Analysis	0.020	P
FSP-1 Irradiation Summary Report	0.027	E
FSP-1 Irradiation Summary Report Complet	0	
FSP-1 Non-Destructive PIE	0.115	P
FSP-1 Destructive PIE	0.117	P
FSP-1 PIE Summary Report	0.242	E
FSP-1 Material Properties Summary Report	0.110	E
FSP-1 Post Irradiation Material Properties	0.067	P
FSP-1 Complete	0.000	

Important to note here is that the consistency ratio (CR) of the project is .15, which over the recommended value of 0.10, but falls under the consistency threshold of 0.20 (Pedrycz & Song, 2011). This is likely due to 15 activities being compared, which is rather large (Forman & Selly, 2001). This issue will be addressed in Section 6.2.1.2.

5.3 Establishing the Learning Baseline

With the learning values defined and the project schedule, the learning baseline is calculated for the project. Again, this is done using the schedule logic and the defined methods for how each activity learns. Given the duration of the project, the project's reporting periods were initially broken into 120 days. This resulted in 30 reporting periods for the project.

Upon review of the schedule, it was determined that some time existed in the middle of the project where the project was awaiting activities to complete in a separate project. Those activities were the production of test articles. The production did not affect the learning of the project, as it was simply a manufacturing step. To display the baseline with respect to work being executed in the project, these six periods were eliminated from the baseline. In those six periods, no resources were used, no money was being spent, no work was being performed, no learning was achieved, no earned value was generated – it was simply a pause and restart. Nothing was occurring in the project. The learning baseline without these no value periods is shown in Figure 5-2. The learning baseline values are shown in Appendix C.

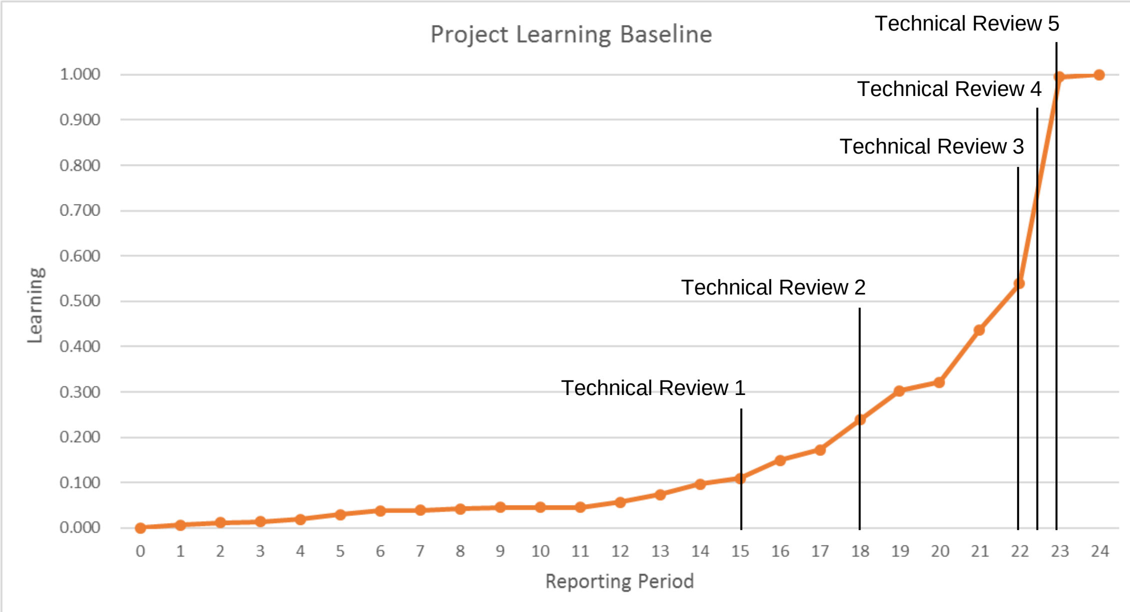

Figure 5-2: Experiment Project's Learning Baseline

5.4 Scheduling the Technical Reviews

The learning baseline establishes recommendations for technical reviews (and the backbone for predicting future values) and is made using the segmented linear regression. In this execution, the segmented linear regressions are done manually, via trial and error. Again, software exists to perform this function, but to meet the needs to regularly monitor the research and development effort's learning progress, it was done manually (Ritzema, 1994). Rather than an optimized statistical solution, segmenting the baseline manually allows management to balance the frequency of reviews with acceptable linear regressions.

Learning begins very slowly in the project and curves sharply at the end. The technical reviews for the project were simple to schedule because of the behavior of the learning. Five technical reviews were scheduled at the breakpoints of 0.109 learning, 0.238 learning, 0.54 learning, 0.7313 learning, and 0.995 learning. An additional review could be done at project completion, but is likely unnecessary as the project has only 0.005% learning left, and any considerations could be covered during project closeout (PMI, 2013.).

Table 5-2 below shows the scheduled technical reviews, the equation of the linear segments, their R-squared values, standard errors of the regressions, and the variable *p-values*.

Table 5-2: Scheduled Technical review and statistical details

Technical Reviews	**1**	**2**	**3**	**4**	**5**
Learning Achievement	0.109	0.238	0.54	0.7313	0.995
Equation	y=0.006x-0.0037	y=0.041x-0.51	y=0.295x-6.266	y=.285x-5.738	y=1.05x-23.23
R-squared	0.899	0.962	0.945	0.913	N/A
Std Error of Regression	0.01	0.013	0.032	0.035	N/A
Variable *p-value*	0	0.019	0.005	0.044	N/A

The first technical review is scheduled to cover up to period 15. This long duration, almost 5 years, covers over half of the project duration. Lengthy as this segment is, the learning achievement marked for the first technical review is 0.109 or just over 10% of the project's uncertainty eliminated. This is little compared to the duration and project spending. Although this is the first scheduled technical review, normal reporting and monitoring of progress can readily be accomplished up to this point (DAU, 2017a).

At 0.238 learning achieved, the second technical review covers the second segment of the project. This segment has a high R^2, and low standard error of the regression, indicating a good fit of the line (Albright et al., 2011). This technical review covers period 15 to period 18 in the baseline schedule, a duration of 360 days. Technical review 3 continues the good fit trend with a high R^2 and good standard error of the regression (Albright et al., 2011). It covers period 18 to period 22, another 480 days.

Technical review 4 is an abnormal case. As shown in the learning baseline, there is no learning value that correlates to 0.7313. Period 22 to period 23 (reminder – these are 120 days), was broken into 30-day reporting periods to increase data points and increase fidelity of the linear segment. This resulted in technical review 4 covering period 22 to period 22.75, only 90 days, but almost 20% of the project learning.

Technical review 5 nearly finishes the project. It was tested to finish the learning and bring that technical review to 1.0. The last two reporting periods only cover 0.005 learning achieved, the linear regression was poor going to 1.0, so the recommendation is for technical review 5 to cover period 22.75 to period 23 - only 30 days. Those happen to be an important 30 days, gaining 0.26 learning at the end of the project. As previously stated, it is recommended to capture the final 0.005 learning at the project closeout.

5.5 Simulating Delayed Achievement

Testing LPM's ability to be predictive of uncertainty required changing the baseline of how the project achieved learning. To simulate this behavior, the project's activities were randomly increased from 0 to 50%. This action would change the behavior of the project's learning and provide data sets to analyze the predictive capability of LPM.

Failure to achieve project learning was not tested in this assessment. That level of anomaly would require management level decisions on acceptable learning performance thresholds. Those decisions would be highly situational.

Table 5-3: New durations for each project activity

Activity Name	Duration	% over	Over Duration
FSP-1 Conceptual Design	186	10	205
FSP-1 FSI Modeling Plan	42	0	42
FSP-1 Conceptual Design Review (30%)	44	20	53
FSP-1 Preliminary Design	215	30	280
FSP-1 60% Design Review	58	50	81
FSP-1 Final Design	174	50	244
FSP-1 Final Design Review	172	10	189
FSP-1 Design Complete	0	30	0
FSP-1 Irradiation Vehicle Fabrication	24	50	34
FSP-1 Flow Testing	192	40	269
FSP-1 Prepare Characterization Plan	69	50	97
FSP-1 Detailed Characterization	222	50	311
FSP-1 Characterization Summary Report	33	40	46
FSP-1 Characterization Complete	0	40	0
FSP-1 Safety Analysis	301	10	331
FSP-1 Pre-Irradiation Material Properties	249	0	249
FSP-1 Integrated Test Assembly	195	30	254
FSP-1 ATR-C Run	42	50	59
FSP-1 Irradiation (Four ATR Cycles)	191	30	248
FSP-1 Cooling and Shipping	88	20	106
FSP-1 As-Run Analysis	44	10	48
FSP-1 Irradiation Summary Report	54	30	70
FSP-1 Irradiation Summary Report Complete	0	20	0
FSP-1 Non-Destructive PIE	110	10	121
FSP-1 Destructive PIE	132	20	158
FSP-1 PIE Summary Report	66	10	73
FSP-1 Material Properties Summary Report	88	0	88
FSP-1 Post Irradiation Material Properties	154	20	185
FSP-1 Complete	0	0	

The table above (Table 5-3) shows the random (via Excel function) assignment of the increase in duration and the resulting new duration. These durations may look inconsistent with time scale of the projects, but those durations only represent the working days. These durations were entered into scheduling software with the applicable project data which schedule the calendar duration of the projects. That action was planned for non-working days which expanded the time-scale.

For this data collection, the assumption was made that management did not take actions to correct these schedule issues. Using the new durations, the learning values are calculated as before, on the delayed schedule. The new project learning achievements by period are shown in Appendix C.

5.6 Predicted Future Learning Value

These new learning values were used to see if the LPM predicted future value was within the uncertainty bounds of the prediction interval. This was accomplished using the process described in Section 4.1.5. The *expected period* calculations developed using the achieved learning from the delayed schedule calculations. Those *expected periods* were advanced to the next reporting period and entered in the linear equation for that segment to solve for the *predicted learning.*

Expectedly, the learning achieved values do not align precisely with the planned technical review values. The technical review values serve as the breakpoints, and therefore are indictors for when to stop using the previous segment's linear equation and pick up the next. As the learning achieved values approached the breakpoints, both the current linear equation and next linear equation were used to calculate *predicted learning.* Further, once the breakpoint value was achieved, an additional point of data was calculated.

Table 5-4 below shows the data set calculated for the project. While calculating the learning values and *predicted learning*, it was necessary to decompose period 24 into quarters. This action was consistent with the treatment of the initial learning baseline, which also decomposed the period at approximately 0.5 learning.

For each period, Table 5-4 is displaying various information. In the first column, *Period*, it is simply stating what the reporting period is. The second column, *Learning Achieved,* is displaying the amount of learning that was achieved during the simulated project. The third column, and the subsequent columns beginning with *Technical Review*, show on what reporting period the previous learning achieved values would fall on the baseline project. They are titled *Technical Review X*, because the linear equations from the titled segments were used to calculate that status. Using status in the *Technical Review X*, the next column, *Prediction*, displays the predicted learning value for each period.

Table 5-4: Learning Values and Predictions

Period	Learning Achieved	Technical Review 1	Prediction	Technical Review 2	Prediction	Technical Review 3	Prediction	Technical Review 4	Prediction	Technical Review 5	Prediction
0	0.000	0.000									
1	0.005	1.450	0.002								
2	0.011	2.450	0.011								
3	0.013	2.783	0.017								
4	0.016	3.283	0.019								
5	0.023	4.450	0.022								
6	0.031	5.783	0.029								
7	0.037	6.783	0.037								
8	0.039	7.117	0.043								
9	0.040	7.283	0.045								
10	0.043	7.783	0.046								
11	0.045	8.117	0.049								
12	0.046	8.283	0.051								
13	0.046	8.283	0.052								
14	0.054	9.617	0.052								
15	0.074	12.950	0.060								
16	0.094	16.283	0.080	14.732							
17	0.150	25.617	0.100	16.098	0.135						
18	0.160		0.156	16.341	0.191						
19	0.211			17.585	0.201						
20	0.263			18.854	0.252	18.500					
21	0.310			20.000	0.304	19.135	0.337				
22	0.342				0.351	19.568	0.384				
23	0.403					20.392	0.416				
24	0.509					21.824	0.477				
24.25	0.531					22.128	0.528	21.998			
24.5	0.554					22.438	0.550	22.079	0.603		
24.75	0.687						0.573	22.545	0.626		
25	0.710							22.625	0.759	22.800	
26	1.000							23.642	0.995	23.076	1.760

6 Results

6.1 Overview

Generating analysis of the experiment is broken into two sections. The first one deals with observations of implementing LPM on a real research and development effort. Those observations are challenges and mitigations in implementing the methodology. The second section will be an assessment of the predictive capability of LPM. Insights and quantitative analysis were completed to support these conclusions.

6.2 Results of Implementing Learning Performance Management

Generating the pairwise comparisons based on the project activities was a relatively simple task that was complicated by issues. First, confusion over the intention of technical uncertainty changed how the process was described. Next, the pairwise comparisons in the project affected the consistency ratio. Finally, sharp learning achievement increases in the project presented challenges.

6.2.1.1 Confusion over terminology

When discussing the comparisons with the technical program manager, the original request was to compare the relative "technical performance risk" of the activities. That jargon, discussed in Section 2.3, was confusing to explain. During discussion, the first clarification was to refine technical risk as "technical uncertainty," also as described in Section 2.3. However, using the term "uncertainty" made an unintended connection to "schedule uncertainty," as used in project management.

Ultimately, the best explanation of the comparison, without confusion was "how much is expected to be learned" toward the technology maturation. Confusing language, and jargon were eliminated, and the comparison was clarified. Learning, and elimination of the unknown, toward the technical goal, was chosen to describe the methodology.

6.2.1.2 Pairwise Comparisons

As mentioned earlier, the consistency ratio for the project was over the recommended value of 0.10 (Saaty, 1986). The value 0.15 indicates that there is consistency in the assessments and they are not random. Randomness would be indicated by a consistency ratio approaching 1.00. Saaty does state that a consistency ratio under 0.2 is tolerable (Pedrycz & Song, 2011). The project does fall within the tolerable level.

Actions could be taken to decrease the consistency ratio. First, breaking the activities into smaller groups to perform pairwise comparisons, then comparing the groups themselves could make the assessments more manageable (Forman & Selly, 2001). The project does have interim milestones. Distilling the activities into groups based on those milestones could have supported smaller grouping. Additionally, consistent with group decision making research, executing the pairwise comparisons by a team of diverse and independent reviews could also improve the result (Lehrer, 2009).

Additionally, developing the pairwise comparisons was arduous. Given the values were calculated manually, and the assessments were garnered on a spreadsheet, the calculation of the learning values took significant effort. Had the number of activities been greater, this process would likely take too much effort to use.

Further research needs to be accomplished in two avenues to overcome this problem. First, by using another multi-criteria decision-making process, or another value assignment method, LPM’s level of effort could be reduced. Second, integrating pairwise comparisons into scheduling software for easy calculation and capturing of assessments should be researched. Software solutions currently exist to capture assessments (i.e. ExpertChoice ©), so the need is to research integration techniques (Bible & Bevins, 2011).

6.2.1.3 Learning Achievement Growth

As discussed previously, the way in which the project achieves learning is problematic. It experiences long periods of slow achievement, followed by sharp increases in growth of learning. The project experiences a long duration of slow growth with 10% of the learning in the first 15 periods. This is followed a large 45% increase of achievement in 120 days during the 23rd period.

While these sharp increases and slow periods cause problems in predicting future values, they also are impractical from a management perspective. Gaining much value in a short period, particularly at the end, forces risk acceptance and investment until the end of the project.

Consistent with other project management processes, risk management processes, and other management processes, it may be advisable to change the schedule logic or iterate planning processes to make the learning achievement more constant. This may make investment and learning more correlated.

6.3 Development of Learning Values

Hypothesis 1 was focused on whether activity-based uncertainty values could be developed using the principles of Risk Value Method. Table 5-1 shows the outcome of developing those learning values. Those learning values were the amount of uncertainty expected to be eliminated, toward the research and development goal. These values were no absolute or exhaustive, but the relative weights of each activity in the project.

Due to the constraints of the research it was not possible to track, the learning values in real time to assess the learning values against the research and development performance. As discussed, the data set came from a project, yet to be completed, nearly 10 years in duration. That project had abandoned EVMS and never implemented TPM, so there was no existing data to compare.

However, distilling the project data exposed information that contributed to the hypothesis. First, as shown in Table 6-1, there is a clear disconnect in the cost of each activity and the impact to learning. This highlights the disconnect in the EVMS method of tracking project progress and progress toward the project goal. At its most egregious, the project's largest learning activity, with almost 25% of the learning, only uses 1% of the project budget. The correlation of learning weight to costs is 0.41, shown in Table 6-1. That value represents a low correlation. A regression performed on these values, shows the R^2 is only 0.17.

Similarly, but not shown, the project's critical path is not aligned with its learning values. Specifically, only 0.413 learning is achieved on the critical path. This means that the project's scheduled activities critical to its completion are disconnected with the majority of learning to occur in the project. Those critical path activities represent 73% of the project budget – a full 30% difference from the learning value. This seems to confirm both the problem with using EVMS to track technical progress and shows that LPM can be used in lieu of EVMS.

By using LPM, decision makers are able to analyze the relative cost benefit of activities based on the expected contribution to learning. Doing so may help make decisions on when to apply resources to activities in order to return them to their baseline schedule, what activities may be spread out, and importantly, show value in context of learning achieved instead of dollars spent.

Table 6-1: Weight to Cost per activity comparison

Activity Name	Total Cost	Percentage of Project Budget	Learning	Percentage of Learning
FSP-1 Conceptual Design	$371,658	6%	0.013	1.30%
FSP-1 FSI Modeling Plan	$30,000	0%	0	0.00%
FSP-1 Conceptual Design Review (30%)	$38,021	1%	0	0.00%
FSP-1 Preliminary Design	$324,417	5%	0.012	1.16%
FSP-1 60% Design Review	$133,163	2%	0	0.00%
FSP-1 Final Design	$136,346	2%	0.007	0.71%
FSP-1 Final Design Review	$81,222	1%	0	0.00%
FSP-1 Design Complete	$0	0%	0	0.00%
FSP-1 Irradiation Vehicle Fabrication	$152,894	2%	0	0.00%
FSP-1 Flow Testing	$300,724	5%	0.014	1.36%
FSP-1 Prepare Characterization Plan	$45,000	1%	0	0.00%
FSP-1 Detailed Characterization	$684,120	11%	0.033	3.28%
FSP-1 Characterization Summary Report	$52,000	1%	0.038	3.78%
FSP-1 Characterization Complete	$0	0%	0	0.00%
FSP-1 Safety Analysis	$370,529	6%	0	0.00%
FSP-1 Pre-Irradiation Material Properties	$307,809	5%	0.034	3.35%
FSP-1 Integrated Test Assembly	$91,578	1%	0	0.00%
FSP-1 ATR-C Run	$105,728	2%	0	0.00%
FSP-1 Irradiation (Four ATR Cycles)	$1,344,559	21%	0.152	15.19%
FSP-1 Cooling and Shipping	$267,136	4%	0	0.00%
FSP-1 As-Run Analysis	$105,780	2%	0.020	2.04%
FSP-1 Irradiation Summary Report	$60,000	1%	0.027	2.74%
FSP-1 Irradiation Summary Report Complet	$0	0%	0	0.00%
FSP-1 Non-Destructive PIE	$366,220	6%	0.115	11.48%
FSP-1 Destructive PIE	$579,064	9%	0.117	11.74%
FSP-1 PIE Summary Report	$60,711	1%	0.242	24.20%
FSP-1 Material Properties Summary Report	$30,017	0%	0.110	10.97%
FSP-1 Post Irradiation Material Properties	$395,575	6%	0.067	6.70%
FSP-1 Complete	$0	0%	0.000	0.00%

The stark difference in how learning is achieved for the project and traditional earned value is shown below (Figure 6-1). In this figure, the baseline earned value for the projected (calculated with the same earning rules as the learning achieved) is compared to the learning values, in percentages. The learning values for each reporting period were simply formatted in percentages, while the earned value was divided by the project total earned value, and represented as a percentage.

As can be seen in the figure, the earned value of the project increases much faster and at a more linear appearing fashion. By period 5, the earned value was 13.7% (in blue) of the project budget, while the learning achieved was only 3.0% (in orange). The gap between learning and earned value continues to grow, as shown at each five intervals. By period 22, the final data point shown below, earned value would say the project had earned 89.6% of its value, whereas only 54.0% of learning was achieved.

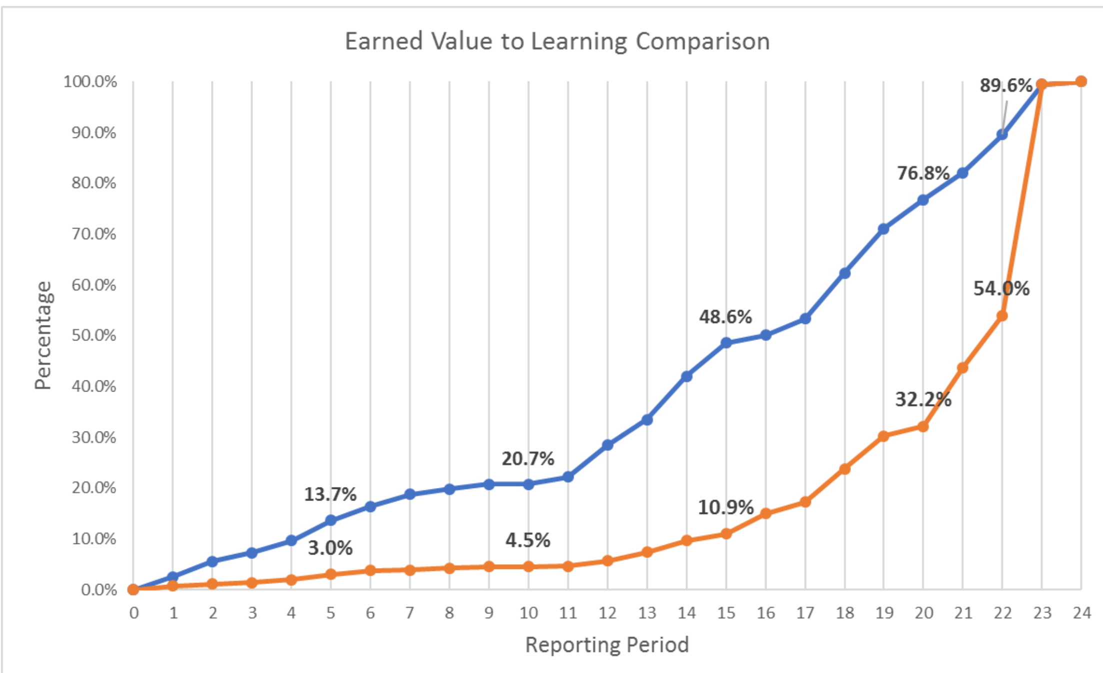

Figure 6-1: Earned Value to Learning Comparison

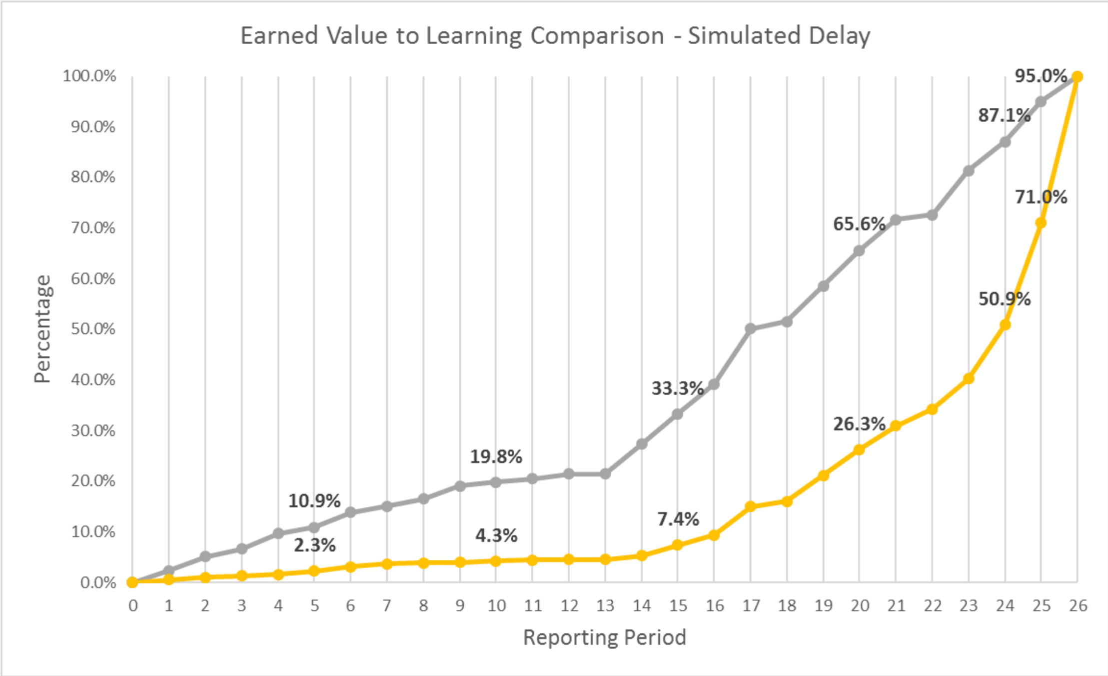

Figure 6-2: Earned Value to Learning Comparison under simulated delay

Figure 6-2 continues to highlight these issues, showing the percentage comparisons of earned value (in grey) and learning achieved (in yellow) for the project under the simulated delays. Each five reporting periods are labeled, showing that the learning achieved slowly grows at a rate slower than the earned value. The earned value is largely disconnected from the learning achieved, especially in the bottom of the convex curve of the learning achieved curve.

The dissimilar behavior of earned value and learning is understandable, given both the schedule logic and how the activities are credited. The learning achievement rapidly increases at the end of project, as it begins to meet its technical goals. The earned value grows as activities are completed – regardless of those activities contribution to the technical performance.

Given this disconnect, LPM performance monitoring is clearly not similar to EVMS.

6.4 LPM Predictive Capability

Hypothesis 2 was whether LPM's predictive function would fall within the calculated 95% prediction interval. To answer this, *predicted learning* values were evaluated to see if they fell within the prediction interval. Three pieces of information were considered for the accuracy, the variance, the percentage of variance, and if the *predicted learning* were within the 95% prediction interval of the regression.

The variance from the realized learning achieved value was calculated by subtracting the *predicted learning* from the actual value. This information gauged the size of variance from the prediction. When the learning values are low, this size is low. Therefore, dividing the variance by the actual value garners a percentage of variance. This percentage puts the amount of variance in perspective with the size of the learning values.

Finally, a determination of whether the actual learning values fall within the 95% prediction intervals of the *predicted learning* values is made. The *predicted learning* values represent data points on the linear regression. If the actual values fall within the prediction intervals, based on standard errors of the estimate of the *predicted values*, the regression should have an acceptable level of uncertainty (Leininger, 2013). Based on how well the data falls within the predictive intervals, the research hypothesis can be confirmed or rejected.

6.4.1 Prediction Interval Results

LPM's predictions for the simulated project prove to be accurate within the bounds of the prediction interval. This result is less about validating accuracy of the prediction, but showing that the forecast method can be used to bound the expectations of those using the predictions. The prediction intervals expand and contract with the standard error of the estimate, which in turn decreases as regression accuracy increases (Leininger, 2013).

For the segment leading to Technical Review 1, all learning achieved values are within all the prediction intervals. The standard error of the regression for this segment was 0.01. Table 6-2 below displays this data. The *learning achieved* column is the actual learning value realized by the simulated project for that period. The *period calculated* column is the period that the previous column's value corresponds to in the segmented linear regression. The *prediction* column is the *predicted learning* for that period. The *variance* and *percentage* values are the difference of the prediction from the actual. The *Lower Bound* and *Upper Bound* columns show the boundaries of the prediction interval, while the *Within* column answers whether the actual learning value falls in that boundary.

Table 6-2: Technical Review 1 Prediction Results

Technical Review 1								
Period	**Learning Achieved**	**Period Calculated**	**Prediction**	**Variance**	**Percentage**	**Lower Bound**	**Upper Bound**	**Within?**
0	0.000	0.000	0.000					
1	0.005	1.450	0.005	0.000	0.0%	(0.017)	0.027	Y
2	0.011	2.450	0.011	0.000	0.0%	(0.011)	0.033	Y
3	0.013	2.783	0.017	(0.004)	-30.8%	(0.005)	0.039	Y
4	0.016	3.283	0.019	(0.003)	-18.8%	(0.003)	0.041	Y
5	0.023	4.450	0.022	0.001	4.3%	(0.000)	0.044	Y
6	0.031	5.783	0.029	0.002	6.5%	0.007	0.051	Y
7	0.037	6.783	0.037	0.000	0.0%	0.015	0.059	Y
8	0.039	7.117	0.043	(0.004)	-10.3%	0.021	0.065	Y
9	0.040	7.283	0.045	(0.005)	-12.5%	0.023	0.067	Y
10	0.043	7.783	0.046	(0.003)	-7.0%	0.024	0.068	Y
11	0.045	8.117	0.049	(0.004)	-8.9%	0.027	0.071	Y
12	0.046	8.283	0.051	(0.005)	-10.9%	0.029	0.073	Y
13	0.046	8.283	0.052	(0.006)	-13.0%	0.030	0.074	Y
14	0.054	9.617	0.052	0.002	3.7%	0.030	0.074	Y
15	0.074	12.950	0.060	0.014	18.9%	0.038	0.082	Y
16	0.094	16.283	0.080	0.014	14.9%	0.058	0.102	Y
17	0.150	25.617	0.100	0.050	33.3%	0.078	0.122	Y
18	0.160		0.156	0.004	2.5%			

The segment leading up to Technical Review 2 has a standard error of the regression of 0.025. For all values leading to Technical Review 2, the learning achieved values fall within the prediction interval. Table 6-3 below displays this data:

Table 6-3: Technical Review 2 Prediction Results

Technical Review 2								
Period	**Learning Achieved**	**Period Calculated**	**Prediction**	**Variance**	**Percentage**	**Lower Bound**	**Upper Bound**	**Within?**
16	0.094	14.732						
17	0.150	16.098	0.135	0.015	10.0%	0.067	0.203	Y
18	0.160	16.341	0.191	(0.031)	-19.4%	0.128	0.254	Y
19	0.211	17.585	0.201	0.010	4.7%	0.138	0.264	Y
20	0.263	18.854	0.252	0.011	4.2%	0.188	0.316	Y
21	0.310	20.000	0.304	0.006	1.9%	0.233	0.375	Y
22	0.342		0.351	(0.009)	-2.6%	0.271	0.431	

Table 6-4: Technical Review 3 Prediction Results

Technical Review 3								
Period	**Learning Achieved**	**Period Calculated**	**Prediction**	**Variance**	**Percentage**	**Lower Bound**	**Upper Bound**	**Within?**
20	0.263	18.500						
21	0.310	19.135	0.337	(0.027)	-8.7%	0.290	0.384	Y
22	0.342	19.568	0.384	(0.042)	-12.3%	0.337	0.431	Y
23	0.403	20.392	0.416	(0.013)	-3.2%	0.370	0.462	Y
24	0.509	21.824	0.477	0.032	6.3%	0.431	0.523	Y
24.25	0.531	22.128	0.528	0.004	0.7%	0.480	0.575	Y
24.5	0.554	22.438	0.550	0.004	0.8%	0.502	0.598	Y
24.75	0.687		0.573	0.114	16.6%			

Table 6-4 above shows the *predicted learning* values for the segment leading to Technical Review 3. Consistent with the learning baseline, reporting period 24 was decomposed into quarters. These additional data points were beneficial in predicting the lead up to the Technical Review, which occurred at 0.54 learning achieved. The standard error of the regression for this segment is 0.032. This represents a looser bound than previous segments. All the *learning achieved* values fall within the prediction interval.

Table 6-5: Technical Review 4 Prediction Results

Technical Review 4								
Period	Learning Achieved	Period Calculated	Prediction	Variance	Percentage	Lower Bound	Upper Bound	Within?
24.25	0.531	21.998						
24.5	0.554	22.079	0.603	(0.048)	-8.7%	0.537	0.669	Y
24.75	0.687	22.545	0.626	0.062	9.0%	0.561	0.690	Y
25	0.710	22.625	0.759	(0.049)	-6.8%	0.695	0.822	Y
26	1.000	23.642	0.995	0.005	0.5%	0.931	1.059	Y

The segment leading to Technical Review 4 had few number of data points and the prediction interval was larger. The standard error of the regression was also larger, at 0.035. All the *predicted learning* values falling within the prediction intervals, as shown in Table 6-5.

All the *predicted learning* values fall within the 95% prediction intervals for individual segments. This would appear to confirm that, in this simulated environment, the methodology is effective to predict future values within the bounds of the prediction interval uncertainty. The lower bound of the prediction intervals does not seem particularly valuable. All but two of the prediction interval's lower bounds are less than the previously achieved learning.

The prediction intervals used the t-distribution, to calculate their values. This was used in lieu of the normal *Z* values intentionally. While the x values, or the reporting period, were characterized, the learning achieved values were unknown. This would indicate that the sample mean was known, but the population mean was unknown (Albright et al., 2011). Additionally, the regression's utilized a small number of data points (less than 50), and therefore would use the t-distribution (Eftekari, 2015).

6.4.2 Validating Data Set

A possible concern that emerged with the research is that the learning achieved values from the simulation are part of the data set, and that predictive intervals are not informative. The concern is the simulation was unknowingly biased in favor of the method by being within the regression.

To test whether this concern is valid, 95% percent confidence intervals were calculated in lieu of the prediction intervals. If 95% of the *predicted learning* values fall into these bounds, it would appear that they were simply part of the data set. None of the segments, however, passed that test, as none reached 95% of *predicted learning* values within bounds.

6.5 Comparison to Forecasting Methods

In order to add value, LPM's forecasting should to be more effective than existing methods. Since LPM is a new method, the existing forecasting methods for EVMS and TPM cannot be directly applied to LPM, although further research is required to determine whether they could be used. However, basic mathematical forecasting methods including linear forecasting, exponential forecasting, and exponential smoothing can be evaluated against LPM using accepted measurements of forecasting accuracy.

Using Excel ©, linear forecasts, exponential forecasts, and exponential smoothing forecasts were made for the learning values. This was done across the entire 26 periods of the project (after it had been delayed). For the exponential smoothing method, the smoothing constant was manually refined to its most accurate – 0.001 (Albright et al., 2013).

For each measurement of forecasting accuracy, the predictions of the methods were compared to the realized learning values from the simulated schedule. The predicted learning values were used from one linear segment, until it reached its baseline technical review value, then the prediction used the next linear segment. As shown in Table 6-6, the LPM prediction shows the least MAD, and therefore performs the best according to this test (Hocking, 2013, p. 210).

MAPE is the comparison of percentages (Montano Moreno et al., 2013). For this measurement, the learning values were on the same scale, so this measurement may not deviate from MAD. As seen in Table 6-6, similar to MAD, LPM performs substantially better, with the least percentage error (Montano Moreno et al., 2013). LPM performs best as well for RMSE and MSE (Thompson, 1990; Willmott & Matsuura, 2005). Table 6-6 shows that LPM performs the significantly better than all the basic forecasting techniques.

Table 6-6: Forecast Performance Comparison

	LPM	Linear Forecast	Exponential Forecast	Exponential Smoothing (.001)
MAD	**0.014**	**0.086**	**0.078**	**0.041**
MAPE	**0.108**	**0.279**	**0.249**	**0.188**
MSE	**0.00043**	**0.02667**	**0.02325**	**0.00635**
RMSE	**0.020**	**0.163**	**0.152**	**0.080**

7 Conclusion

7.1 Experiment

The experiment showed the ability to implement the proposed Learning Performance Measurement methodology described in Chapter 4. The new method provided the ability to calculate learning weights and monitor the project status throughout the duration of the project. Predicted learning measurements were made based on a delayed schedule. Those predicted learning values were verified as being able to predict future values within the 95% confidence.

The experimental results discussed in Chapter 6 provide evidence that all three hypotheses can be accepted:

H_1: An activity-based uncertainty measurement can be developed based on Risk Value Method concepts.

H_2: Learning Performance Measurement predictions are valid within the 95% prediction interval.

H_3: Learning Performance Measurement predictions perform better than mathematical standard forecasting methods.

By accepting these hypotheses, the research question is answered positively.

7.2 Limitations

7.2.1 Additional Progress and Project Metrics

As defined in the methodology, the LPM process is focused on managing the technical progress of the R&D effort. The statuses provided by the method give project managers and leadership insight into how the activity is progressing toward its technical goal. Opposite from EVMS, LPM does not provide insight into the project status, in terms of cost or schedule when the project is not progressing toward its technical goals (Abdullah, Hamzah, Ismail & Razak, n.d.; Bower & Finegan, 2009).

This is seen in the experiment described in Section 5 and further displayed in the discussion in Section 6.3. Specifically, there are many activities that do not contribute to learning, but expend resources. The execution of these resources will still need assessment and tracking, consistent with project monitoring (Meredith & Mantel, 2015). Their impact on the schedule needs tracked for how the project is progressing towards its plan (Meredith & Mantel, 2015). Due to these factors, it is clear that LPM is not a standalone metric, but should be an element in a suite of metrics used to track statuses.

7.2.2 Subjectivity

While LPM begins to quantify technical progress, there is still significant subjectivity in the method. AHP has been criticized with various limitations and issues identified in its methodology (Gass, 2015). Early stage R&D is inherently uncertain and involves discovery of the previously unknown (Browning et al, 2002; OECD, 2015). The LPM attempts to begin bounding those subjective elements, but it should be acknowledged that much subjectivity still exists due to the project characteristics and methodology.

7.3 Further Research

Numerous opportunities exist for future research. Primarily, LPM should continue to be applied to projects, programs, and systems to further evaluate and refine the method, associated documents, and tools. Particularly, system-level implementations of LPM to analyze performance and shortcomings are needed. As previously noted, further research is required to reduce the level of effort to develop the learning values.

Further research is also needed into the use of predictive intervals as tolerance bands in LPM. Since this research doesn't suppose a method to calculate tolerance bands, there are no recommendations for how to make them. As TPM, and therefore LPM, are used in risk management, using the calculated predictive intervals could both set expectations for management, as well provide a basis to simply assess the uncertainty of the prediction (Sears & Taylor, 1984). This simple usage was shown by Savelli in 2013 when she observed that "the majority of participants using predictive intervals were able to identify forecasts with greater uncertainty despite the fact that the relationship between interval width and uncertainty had not been explained to them."

Additionally, there are possibilities to intersect LPM and TPM measurements to provide system level information for systems with varying technology maturities. This research would focus on integration of learning and quantified technical performance.

The forecasting method developed for LPM is very elementary. Further research should be performed to tailor a forecasting method to LPM equities, as have been done for EVMS and TPM (Lipke et al., 2009; Batsailer & Vanhoucke, 2017; Batsailer & Vanhoucke, 2015; Eggstaff et al., 2014). This research should keep in mind the dual goal of the LPM regressions – setting technical review milestones and developing expectation management of learning achievement for decision makers.

Lastly, research should be accomplished to assess the utility of making the linear segments a cost/benefit method. Research into regressing the cost and learning together, into linear segments, so that decision makers can balance investment and learning together. Such a concept could support various go / no-go methods, such as Stage-Gate (Cooper, 2008).

References

- Oxford English Dictionary.- *"Research and development, n."*.

A. D'Amico, B. O'Brien, & M. Larkin. (2013). Building a bridge across the transition chasm. *IEEE Security & Privacy, 11*(2), 24-33. doi:10.1109/MSP.2012.160

Abdullah, A. R. R. B., Hamzah, A. R., Ismail, M. Y., & Razak, A. R. A.Integration of quality measure in project control system. *International Journal of Construction Project Management, 7*(1), 57-75. Retrieved from http://search.proquest.com.proxygw.wrlc.org/docview/1695165145?accountid=11243

Adams, R., Bessant, J., & Phelps, R. (2006). Innovation management measurement: A review. *International Journal of Management Reviews, 8*(1), 21-47. doi:10.1111/j.1468-2370.2006.00119.x

Albright, S. C., Winston, W. L., & Zappe, Christopher J. (Christopher James). (2011). *Data analysis and decision making with microsoft excel* (Third Edition ed.). Pacific Grove, CA ; London: Brooks/Cole-Thomson Learning.

Anbari, F. T. (2003). Earned value project management method and extensions. *Project Management Journal, 34*(4), 12-23. Retrieved from http://proxygw.wrlc.org/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=bth&AN=11583963&site=ehost-live

Azizian, N., Mazzuchi, T., Sarkani, S., & Rico, D. F. (2011). A framework for evaluating technology readiness, system quality, and program performance of U.S. DoD acquisitions. *Systems Engineering, 14*(4), 410-426. doi:10.1002/sys.20186

Batselier, J., & Vanhoucke, M. (2015). Evaluation of deterministic state-of-the-art forecasting approaches for project duration based on earned value management. *International Journal of Project Management, 33*(7), 1588-1596. doi:http://dx.doi.org.proxygw.wrlc.org/10.1016/j.ijproman.2015.04.003

Batselier, J., & Vanhoucke, M. (2017). Improving project forecast accuracy by integrating earned value management with exponential smoothing and reference class forecasting. *International Journal of Project Management, 35*(1), 28-43. doi:https://doi-org.proxygw.wrlc.org/10.1016/j.ijproman.2016.10.003

Bible, M. J., & Bivins, S. S. (2011). *Mastering project portfolio management*. Fort Lauderdale, FL: J. Ross Publishing.

Bower, D. C., & Finegan, A. D. (2009). New approaches in project performance evaluation techniques. *Int J Managing Projects in Bus, 2*(3), 435-444. doi:10.1108/17538370910971072

Bunruamkaew, K. (2012). *How to do AHP analysis in excel*

Cannon, N. J., Ulferts, G. W., & Howard, T. L.Research and development investment. *Journal of Business & Economics Research (Online), 12*(3), 291-n/a. Retrieved from http://search.proquest.com.proxygw.wrlc.org/docview/1665183643?accountid=11243

Cleland, D. I. (2004). *Field guide to project management*. Hoboken, N.J.: J. Wiley.

Cooper, R. G. (2008). Perspective: The stage-gate® idea-to-launch Process—Update, what's new, and NexGen systems*. *Journal of Product Innovation Management, 25*(3), 213-232. doi:10.1111/j.1540-5885.2008.00296.x

Crawford, L. H., & Helm, J. (2009). Government and governance: The value of project management in the public sector. *Project Management Journal, 40*(1), 73-87. doi:10.1002/pmj.20107

DAU. (2017). Chapter 3 systems engineering. Retrieved from https://www.dau.mil/tools/dag/Pages/DAG-Page-Viewer.aspx?source=https://www.dau.mil/guidebooks/Shared%20Documents%20HTML/Chapter%203%20Systems%20Engineering.aspx

DAU. (3/1/2017). **Technical performance measurement (TPM)** Retrieved from https://dap.dau.mil/acquipedia/Pages/ArticleDetails.aspx?aid=7c1d9528-4a9e-4c3a-8f9e-6e0ff93b6ccb

Davenport, T. H. (2013). *The analytics advantage.* (Report). United Kingdom: Deloitte Touche Tohmatsu Limited.

De Gooijer, J. G., & Hyndman, R. J. (2006). *25 years of time series forecasting* doi:http://dx.doi.org.proxygw.wrlc.org/10.1016/j.ijforecast.2006.01.001

Defense Science Board. (2012). *Report of the defense science board task force on basic research.* (Report). Washington, DC: Office of the Under Secretary of Defense for Acquisition, Technology and Logistics.

Dekking, M., Kraaikamp, C., Lophaa, H.P., Meester, L.E., (2005). *Modern introduction to probability and statistics : Understanding why and how*. London: Springer.

Dustin, G., Bharat, M., & Jitendra, M. (2014). Competitive advantage and motivating innovation. *Advances in Management, 7*(1), 1-7. Retrieved from http://search.proquest.com.proxygw.wrlc.org/docview/1491286193?accountid=11243

Eckhause, J. M., Hughes, D. R., & Gabriel, S. A. (2009). Evaluating real options for mitigating technical risk in public sector R&D acquisitions. *International Journal of Project Management*, *27*(4), 365-377. doi:https://doi-org.proxygw.wrlc.org/10.1016/j.ijproman.2008.05.015

Eftekari, R. (2015). *Confidence interval and hypothesis testing - lecture 7*. Washington, DC:

Eggstaff, J. W., Mazzuchi, T. A., & Sarkani, S. (2014). The development of progress plans using a performance-based expert judgment model to assess technical performance and risk. *Systems Engineering*, *17*(4), 375-391. doi:10.1002/sys.21273

Fazli, S. (2013). Risks identification and ranking using AHP and group decision making technique: Presenting "R index". *Management Science Letters*, *3*(2), 613; 613-624; 624.

Forman, E. H., Selly, M. A., & ProQuest (Firm). (2001). *Decision by objectives [electronic resource] : How to convince others that you are right*. River Edge, N. J.: World Scientific.

GAO. (1999). *BEST PRACTICES: Better management of technology development can improve weapon system outcomes.* (No. GAO/NSIAD-99-162). Washington, DC: US Government Accountability Office.

GAO. (2012). *DEFENSE ACQUISITIONS: Assessments of selected weapon programs.* (No. GAO-12-400SP). Washington, DC: US Government Accountability Office.

GAO. (2016). *Technology readiness assessment guide: Best practices for evaluating the readiness of technology for use in acquisition programs and projects.* (No. GAO-16-410G). Washington, DC: US Government Accountability Office.

Garvey, P. R., & Cho, C. (2005). *An index to measure and monitor a system-of-systems' performance risk.* (No. 05-0144). Mclean, VA: MITRE Corporation.

Gass, S. I. (2005). Model world: The great debate-MAUT versus AHP. Interfaces, 35(4), 308-312. Retrieved from https://search-proquest-com.proxygw.wrlc.org/docview/217112431?accountid=11243

Godin, B. (2006). The linear model of innovation: The historical construction of an analytical framework. *Science, Technology, & Human Values*, *31*(6), 639-667. Retrieved from http://www.jstor.org.proxygw.wrlc.org/stable/29733964

Godin, B. (2015). Models of innovation: Why models of innovation are models, or what work is being done in calling them models? *Soc Stud Sci, 45*(4), 570-596. doi:10.1177/0306312715596852

H.P.Ritzema (Ed.). (1994). 6:Frequency and regression analysis of hydrologic data. *Drainage principles and applications, publication 16,* (Second Edition ed.,). Wageningen, The Netherlands: International Institute for Land Reclamation and Improvement (ILRI).

Hair, J. F. (2006). *Multivariate data analysis*. Upper Saddle River, N.J.: Pearson Prentice Hall.

Harker, P. T., & Vargas, L. G.The theory of ratio scale estimation: Saaty's analytic hierarchy process. *Management Science, 33*(11), 1383. Retrieved from http://search.proquest.com.proxygw.wrlc.org/docview/213205986?accountid=11243

Hocking, R. R. (2013). *Methods and applications of linear models : Regression and the analysis of variance*. Somerset: Wiley.

Hong, L. C. (2014). Improving forecasting accuracy of project earned value metrics: Linear modeling approach. *Journal of Management in Engineering, 30*(2), 135-145. doi:10.1061/(ASCE)ME.1943-5479.0000187

Hubbard, D. W. (2009). *Failure of risk management : Why it's broken and how to fix it*. Hoboken, N.J.: Wiley.

Johnson, C. (2006, Apr 2006). Implementing an ANSI/EIA-748-compliant earned value management system. *Contract Management, 46*, 36-38,40-43. Retrieved from http://search.proquest.com.proxygw.wrlc.org/docview/196310430?accountid=11243

Kahn, C., & McGourty, S. (2009). *Performance management at R&D organizations.* (). Mclean, VA: MITRE Corporation.

Kerzner, H., & Saladis, F. P. (2009). *Value-driven project management*. Hoboken, N.J.: Wiley.

Khamooshi, H., & Golafshani, H. (2014). EDM: Earned duration management, a new approach to schedule performance management and measurement. *International Journal of Project Management, 32*(6), 1019-1041. doi:http://dx.doi.org.proxygw.wrlc.org/10.1016/j.ijproman.2013.11.002

Kwak, Y. H., & Anbari, F. T. (2012). History, practices, and future of earned value management in government: Perspectives from NASA. *Project Management Journal, 43*(1), 77-90. doi:10.1002/pmj.20272

Kynn, M. (2008). The 'heuristics and biases' bias in expert elicitation. *Journal of the Royal Statistical Society.Series A (Statistics in Society), 171*(1), 239-264. Retrieved from http://www.jstor.org.proxygw.wrlc.org/stable/30130739

Lehrer, J. (2009). *How we decide*. Boston: Houghton Mifflin Harcourt.

Leininger, T. (2013). *Unit 6: Simple linear regression* [Lecture 3: Confidence and prediction intervals for SLR] Duke University.

Lipke, W., & Henderson, K. (2006). Earned schedule: An emerging enhancement to earned value management. *Crosstalk, 19*(11), 26-30.

Lipke, W., Zwikael, O., Henderson, K., & Anbari, F. (2009). *Prediction of project outcome* doi:http://dx.doi.org.proxygw.wrlc.org/10.1016/j.ijproman.2008.02.009

Locatelli, G., Mancini, M., & Romano, E. (2014). Systems engineering to improve the governance in complex project environments. *International Journal of Project Management, 32*(8), 1395-1410. doi:https://doi-org.proxygw.wrlc.org/10.1016/j.ijproman.2013.10.007

M. B. Privitera, M. Design, & J. Johnson. (2009). Interconnections of basic science research and product development in medical device design. Paper presented at the *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society,* 5595-5598. doi:10.1109/IEMBS.2009.5333492

Mahafza, S., Componation, P., & Tippett, D. (2005). A preformance-based technology assessment methodology to support DOD acquisition. *Defense Acquisition Review Journal, 11*(3), 269-282.

Meredith, J. R., Mantel, S. J., author, & Shafer, S. M., author. (2015). *Project management : A managerial approach*. Hoboken, N.J.: Wiley.

Montaño Moreno, J. J., Palmer Pol, A., Abad, A. S., & Blasco, B. C. (2013). Using the R-MAPE index as a resistant measure of forecast accuracy. *Psicothema, 25*(4), 500-506. doi:10.7334/psicothema2013.23

Moorhouse, D. J.*Detailed definitions and guidance for application of technology readiness levels* - American Institute of Aeronautics and Astronautics. doi:-10.2514/2.2916

NIST/SEMATECH e-handbook of statistical methods. (2012). Retrieved from http://www.itl.nist.gov/div898//handbook/pmd/section5/pmd511.htm

NSF. (1953). *The third annual report of the national science foundation.* (No. 3). Washington, DC: National Science Foundation.

OECD. (2015). *Frascati manual 2015* Organisation for Economic Co-operation and Development. doi:10.1787/9789264239012-en

Project Management Institute. (2013). *Guide to the project management body of knowledge (PMBOK® guide) (5th edition)* Project Management Institute, Inc. (PMI). Retrieved from http://app.knovel.com/hotlink/toc/id:kpGPMBKPM1/guide-project-management/guide-project-management

Roedler, G., & Jones, C. (2005). *Technical measurement.* (No. INCOSE-TP-2003-020-01).INCOSE.

Saaty, T. L. (1986). Axiomatic foundation of the analytic hierarchy process. *Management Science, 32*(7), 841-855. Retrieved from http://www.jstor.org.proxygw.wrlc.org/stable/2631765

Sage, A. P., & Rouse, W. B.3.1 the process of risk assessment and management. *Handbook of systems engineering and management (2nd edition)* () John Wiley & Sons. Retrieved from http://app.knovel.com/hotlink/pdf/id:kt007COZW1/handbook-systems-engineering/process-risk-assessment

Savci, S., & Kayis, B. (2006). Knowledge elicitation for risk mapping in concurrent engineering projects. *International Journal of Production Research, 44*(9), 1739-1755. Retrieved from http://proxygw.wrlc.org/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=bth&AN=20937275&site=ehost-live

Savelli, S. (2013). The advantages of predictive interval forecasts for non-expert users and the impact of visualizations. *Applied Cognitive Psychology, 27*(4), 527; 527-541; 541.

Schilling, M. A. (2013). *Strategic management of technological innovation* (Fourth edition ed.). New York, NY: McGraw-Hill.

Sears, B., & Taylor, E. P. (1984). *Technial performance measurement handbook.* (No. V-4062-05). Fort Belvoir, VA: Defense Systems Management College.

Sharon, A., de Weck, O. L., & Dori, D. (2011). Project management vs. systems engineering management: A practitioners' view on integrating the project and product domains. *Systems Engineering, 14*(4), 427-440. doi:10.1002/sys.20187

Spence, J. R., & Stanley, D. J.Prediction interval: What to expect when you're expecting ... A replication. *PLoS One, 11*(9), n/a. doi:http://dx.doi.org.proxygw.wrlc.org/10.1371/journal.pone.0162874

T. R. Browning. (2014). A quantitative framework for managing project value, risk, and opportunity. *IEEE Transactions on Engineering Management*, *61*(4), 583-598. doi:10.1109/TEM.2014.2326986

T. R. Browning, J. J. Deyst, S. D. Eppinger, & D. E. Whitney. (2002). Adding value in product development by creating information and reducing risk. *IEEE Transactions on Engineering Management, 49*(4), 443-458. doi:10.1109/TEM.2002.806710

Thompson, P. A. (1990). *An MSE statistic for comparing forecast accuracy across series* doi:http://dx.doi.org.proxygw.wrlc.org/10.1016/0169-2070(90)90007-X

Townsend, L. A., Mazzuchi, T. A., & Sarkani, S. (2014). A schedule-performance approach for level-of-effort tasks. *Engineering Management Journal, 26*(1), 21-30. Retrieved from http://search.proquest.com.proxygw.wrlc.org/docview/1512705092?accountid=11243

Vandevoorde, S., & Vanhoucke, M. (2006). *A comparison of different project duration forecasting methods using earned value metrics* doi:http://dx.doi.org.proxygw.wrlc.org/10.1016/j.ijproman.2005.10.004

Volkert, R., Stracener, J., & Yu, J. (2014). Incorporating a measure of uncertainty into systems of systems development performance measures. *Systems Engineering, 17*(3), 297-312. doi:10.1002/sys.21270

W. Pedrycz, & M. Song. (2011). Analytic hierarchy process (AHP) in group decision making and its optimization with an allocation of information granularity. *IEEE Transactions on Fuzzy Systems, 19*(3), 527-539. doi:10.1109/TFUZZ.2011.2116029

Wedley, W. C. (1993). *Consistency prediction for incomplete AHP matrices* doi:http://dx.doi.org/10.1016/0895-7177(93)90183-Y

What is systems engineering?. (2017). Retrieved from http://www.incose.org/AboutSE/WhatIsSE

Willmott, C. J., & Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Research, 30*(1), 79-82. Retrieved from http://www.jstor.org.proxygw.wrlc.org/stable/24869236

Appendix A – GAO Technology Readiness Levels

Technology readiness level (TRL)	Description
1 Basic principles observed and reported	Lowest level of technology readiness. Scientific research begins to be translated into applied research and development. Examples include paper studies of a technology's basic properties.
2 Technology concept and/or application formulated	Invention begins. Once basic principles are observed, practical applications can be invented. Applications are speculative, and there may be no proof or detailed analysis to support the assumptions. Examples are limited to analytic studies.
3 Analytical and experimental critical function and/or characteristic proof of concept	Active research and development is initiated. This includes analytical studies and laboratory studies to physically validate the analytical predictions of separate elements of the technology. Examples include components that are not yet integrated or representative.
4 Component and/or breadboard validation in laboratory environment	Basic technological components are integrated to establish that they will work together. This is relatively low fidelity compared with the eventual system. Examples include integration of ad hoc hardware in the laboratory.
5 Component and/or breadboard validation in relevant environment	Fidelity of breadboard technology increases significantly. The basic technological components are integrated with reasonably realistic supporting elements so they can be tested in a simulated environment. Examples include high fidelity laboratory integration of components.
6 System/subsystem model or prototype demonstration in a relevant environment	Representative model or prototype system, which is well beyond that of TRL 5, is tested in its relevant environment. Represents a major step up in a technology's demonstrated readiness. Examples include testing a prototype in a high-fidelity laboratory environment or in a simulated operational environment.
7 System prototype demonstration in an operational environment	Prototype near or at planned operational system. Represents a major step up from TRL 6 by requirement demonstration of an actual system prototype in an operational environment (e.g., in an aircraft, a vehicle, or space).
8 Actual system completed and qualified through test and demonstration	Technology has been proven to work in its final form and under expected conditions. In almost all cases, this TRL represents the end of true system development. Examples include developmental test and evaluation of the system in its intended weapon system to determine if it meets design specifications.
9 Actual system proven through successful mission operations	Actual application of the technology in its final form and under mission conditions, such as those encountered in operational test and evaluation. Examples include using the system under operational mission conditions.

Figure A-1: Technology Readiness Level Descriptions

(from Technology readiness assessment guide: Best practices for evaluating the readiness of technology for use in acquisition programs and projects. (No. GAO-16-410G). Washington, DC: US Government Accountability Office.)

Appendix B – Pairwise Comparison Outcomes

Amount expected to learn…										
Activity Name	Extremely More	Very Strongly More	Strongly More	More	Even	More	Strongly More	Very Strongly More	Extremely More	Activity Name
FSP-1 Conceptual Design				x						FSP-1 Preliminary Design
FSP-1 Conceptual Design				x						FSP-1 Final Design
FSP-1 Conceptual Design					x					FSP-1 Flow Testing
FSP-1 Conceptual Design							x			FSP-1 Detailed Characterization
FSP-1 Conceptual Design							x			FSP-1 Characterization Summary Report
FSP-1 Conceptual Design						x				FSP-1 Pre-Irradiation Material Properties
FSP-1 Conceptual Design									x	FSP-1 Irradiation (Four ATR Cycles)
FSP-1 Conceptual Design						x				FSP-1 As-Run Analysis
FSP-1 Conceptual Design							x			FSP-1 Irradiation Summary Report
FSP-1 Conceptual Design								x		FSP-1 Non-Destructive PIE
FSP-1 Conceptual Design								x		FSP-1 Destructive PIE
FSP-1 Conceptual Design									x	FSP-1 PIE Summary Report
FSP-1 Conceptual Design								x		FSP-1 Material Properties Summary Report
FSP-1 Conceptual Design								x		FSP-1 Post Irradiation Material Properties
FSP-1 Preliminary Design			x							FSP-1 Final Design
FSP-1 Preliminary Design					x					FSP-1 Flow Testing
FSP-1 Preliminary Design							x			FSP-1 Detailed Characterization
FSP-1 Preliminary Design								x		FSP-1 Characterization Summary Report
FSP-1 Preliminary Design							x			FSP-1 Pre-Irradiation Material Properties
FSP-1 Preliminary Design									x	FSP-1 Irradiation (Four ATR Cycles)
FSP-1 Preliminary Design					x					FSP-1 As-Run Analysis
FSP-1 Preliminary Design							x			FSP-1 Irradiation Summary Report
FSP-1 Preliminary Design								x		FSP-1 Non-Destructive PIE
FSP-1 Preliminary Design								x		FSP-1 Destructive PIE
FSP-1 Preliminary Design									x	FSP-1 PIE Summary Report
FSP-1 Preliminary Design									x	FSP-1 Material Properties Summary Report
FSP-1 Preliminary Design								x		FSP-1 Post Irradiation Material Properties
FSP-1 Final Design							x			FSP-1 Flow Testing
FSP-1 Final Design								x		FSP-1 Detailed Characterization
FSP-1 Final Design									x	FSP-1 Characterization Summary Report
FSP-1 Final Design							x			FSP-1 Pre-Irradiation Material Properties
FSP-1 Final Design									x	FSP-1 Irradiation (Four ATR Cycles)
FSP-1 Final Design							x			FSP-1 As-Run Analysis
FSP-1 Final Design								x		FSP-1 Irradiation Summary Report
FSP-1 Final Design								x		FSP-1 Non-Destructive PIE
FSP-1 Final Design								x		FSP-1 Destructive PIE
FSP-1 Final Design									x	FSP-1 PIE Summary Report
FSP-1 Final Design									x	FSP-1 Material Properties Summary Report
FSP-1 Final Design								x		FSP-1 Post Irradiation Material Properties
FSP-1 Flow Testing					x					FSP-1 Detailed Characterization
FSP-1 Flow Testing						x				FSP-1 Characterization Summary Report
FSP-1 Flow Testing						x				FSP-1 Pre-Irradiation Material Properties
FSP-1 Flow Testing									x	FSP-1 Irradiation (Four ATR Cycles)
FSP-1 Flow Testing							x			FSP-1 As-Run Analysis
FSP-1 Flow Testing								x		FSP-1 Irradiation Summary Report
FSP-1 Flow Testing								x		FSP-1 Non-Destructive PIE
FSP-1 Flow Testing								x		FSP-1 Destructive PIE
FSP-1 Flow Testing									x	FSP-1 PIE Summary Report
FSP-1 Flow Testing								x		FSP-1 Material Properties Summary Report
FSP-1 Flow Testing								x		FSP-1 Post Irradiation Material Properties

Figure B-1: Part 1 of 2 completed pairwise comparison for experiment

FSP-1 Detailed Characterization								x		FSP-1 Characterization Summary Report
FSP-1 Detailed Characterization					x					FSP-1 Pre-Irradiation Material Properties
FSP-1 Detailed Characterization								x		FSP-1 Irradiation (Four ATR Cycles)
FSP-1 Detailed Characterization			x							FSP-1 As-Run Analysis
FSP-1 Detailed Characterization				x						FSP-1 Irradiation Summary Report
FSP-1 Detailed Characterization							x			FSP-1 Non-Destructive PIE
FSP-1 Detailed Characterization							x			FSP-1 Destructive PIE
FSP-1 Detailed Characterization								x		FSP-1 PIE Summary Report
FSP-1 Detailed Characterization								x		FSP-1 Material Properties Summary Report
FSP-1 Detailed Characterization					x					FSP-1 Post Irradiation Material Properties
FSP-1 Characterization Summary Report					x					FSP-1 Pre-Irradiation Material Properties
FSP-1 Characterization Summary Report								x		FSP-1 Irradiation (Four ATR Cycles)
FSP-1 Characterization Summary Report					x					FSP-1 As-Run Analysis
FSP-1 Characterization Summary Report					x					FSP-1 Irradiation Summary Report
FSP-1 Characterization Summary Report							x			FSP-1 Non-Destructive PIE
FSP-1 Characterization Summary Report							x			FSP-1 Destructive PIE
FSP-1 Characterization Summary Report								x		FSP-1 PIE Summary Report
FSP-1 Characterization Summary Report							x			FSP-1 Material Properties Summary Report
FSP-1 Characterization Summary Report						x				FSP-1 Post Irradiation Material Properties
FSP-1 Pre-Irradiation Material Properties	x									FSP-1 ATR-C Run
FSP-1 Pre-Irradiation Material Properties									x	FSP-1 Irradiation (Four ATR Cycles)
FSP-1 Pre-Irradiation Material Properties				x						FSP-1 As-Run Analysis
FSP-1 Pre-Irradiation Material Properties					x					FSP-1 Irradiation Summary Report
FSP-1 Pre-Irradiation Material Properties							x			FSP-1 Non-Destructive PIE
FSP-1 Pre-Irradiation Material Properties							x			FSP-1 Destructive PIE
FSP-1 Pre-Irradiation Material Properties								x		FSP-1 PIE Summary Report
FSP-1 Pre-Irradiation Material Properties								x		FSP-1 Material Properties Summary Report
FSP-1 Pre-Irradiation Material Properties					x					FSP-1 Post Irradiation Material Properties
FSP-1 Irradiation (Four ATR Cycles)	x									FSP-1 As-Run Analysis
FSP-1 Irradiation (Four ATR Cycles)	x									FSP-1 Irradiation Summary Report
FSP-1 Irradiation (Four ATR Cycles)				x						FSP-1 Non-Destructive PIE
FSP-1 Irradiation (Four ATR Cycles)				x						FSP-1 Destructive PIE
FSP-1 Irradiation (Four ATR Cycles)							x			FSP-1 PIE Summary Report
FSP-1 Irradiation (Four ATR Cycles)						x				FSP-1 Material Properties Summary Report
FSP-1 Irradiation (Four ATR Cycles)				x						FSP-1 Post Irradiation Material Properties
FSP-1 As-Run Analysis					x					FSP-1 Irradiation Summary Report
FSP-1 As-Run Analysis								x		FSP-1 Non-Destructive PIE
FSP-1 As-Run Analysis								x		FSP-1 Destructive PIE
FSP-1 As-Run Analysis									x	FSP-1 PIE Summary Report
FSP-1 As-Run Analysis								x		FSP-1 Material Properties Summary Report
FSP-1 As-Run Analysis							x			FSP-1 Post Irradiation Material Properties
FSP-1 Irradiation Summary Report								x		FSP-1 Non-Destructive PIE
FSP-1 Irradiation Summary Report								x		FSP-1 Destructive PIE
FSP-1 Irradiation Summary Report									x	FSP-1 PIE Summary Report
FSP-1 Irradiation Summary Report								x		FSP-1 Material Properties Summary Report
FSP-1 Irradiation Summary Report							x			FSP-1 Post Irradiation Material Properties
FSP-1 Non-Destructive PIE					x					FSP-1 Destructive PIE
FSP-1 Non-Destructive PIE								x		FSP-1 PIE Summary Report
FSP-1 Non-Destructive PIE				x						FSP-1 Material Properties Summary Report
FSP-1 Non-Destructive PIE				x						FSP-1 Post Irradiation Material Properties
FSP-1 Destructive PIE							x			FSP-1 PIE Summary Report
FSP-1 Destructive PIE				x						FSP-1 Material Properties Summary Report
FSP-1 Destructive PIE				x						FSP-1 Post Irradiation Material Properties
FSP-1 PIE Summary Report			x							FSP-1 Material Properties Summary Report
FSP-1 PIE Summary Report				x						FSP-1 Post Irradiation Material Properties
FSP-1 Material Properties Summary Report					x					FSP-1 Post Irradiation Material Properties

Figure B-2: Part 2 of 2 completed pairwise comparison for experiment

Appendix C – Learning Performance by Period

Period	0	1	2	3	4	5	6	7	8	9	10	11	12	13
Learning	0.000	0.005	0.011	0.013	0.016	0.023	0.031	0.037	0.039	0.040	0.043	0.045	0.046	0.046

Period	14	15	16	17	18	19	20	21	22	23	24	25	26
Learning	0.054	0.074	0.094	0.150	0.160	0.211	0.263	0.310	0.342	0.403	0.509	0.710	1.000

Figure C-1: Learning Performance Over Duration

Period	0	1	2	3	4	5	6	7	8	9	10	11	12
Learning	0.000	0.007	0.012	0.014	0.020	0.030	0.038	0.039	0.042	0.045	0.045	0.046	0.057

Period	13	14	15	16	17	18	19	20	21	22	23	24
Learning	0.074	0.097	0.109	0.149	0.173	0.238	0.303	0.322	0.437	0.540	0.995	1.000

Figure C-2: Learning Performance Baseline