An Empirical Longitudinal Analysis of Agile Methodologies and Firm Financial Performance

by Andrew L. Bennett

B.S. in Physics, May 2001, James Madison University
MBA in International Business and Entrepreneurship, December 2008, 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 fulfillment of the requirements
for the degree of Doctor of Engineering

January 10, 2019

Praxis directed by

Amir Etemadi
Assistant Professor of Engineering and Applied Science

The School of Engineering and Applied Science of The George Washington University certifies that Andrew Bennett has passed the Final Examination for the degree of Doctor of Engineering as of October 16, 2018. This is the final and approved form of the praxis.

An Empirical Longitudinal Analysis of Agile Methodologies and Firm Financial Performance

Andrew Bennett

Praxis Research Committee:

Amir Etemadi, Assistant Professor of Engineering and Applied Science, Praxis Director

Timothy Blackburn, Professorial Lecturer of Engineering Management and Systems Engineering, Committee Member

Ebrahim Malalla, Visiting Associate Professor of Engineering and Applied Science, Committee Member

© Copyright 2019 by Andrew L. Bennett
All rights reserved

Acknowledgements

The author would first like to thank two of my initial advisors, Dr. Andreas Garstenaur and Dr. Tim Blackburn for their guidance and support early in my pursuit of a doctorate at George Washington University.

Additional thanks are extended to Dr. Amir Etemadi, my advisor for this Praxis. Without his help, the completion of this Praxis may not have been possible.

Finally, the author wishes to express his most profound gratitude to his wife Dana and children, Samantha and Miles for providing ongoing support and encouragement through this course of study.

Abstract of Praxis

An Empirical Longitudinal Analysis of Agile Methodologies and Firm Financial Performance

Agile Software Development methods such as Scrum, SAFe, Kanban, and Large Scale Agile (LeSS) promise substantial benefits in terms of productivity, customer satisfaction, employee satisfaction, quality project management overhead, and time to market. As Agile methods have become widespread in the software development industry and begin to take root in the overall business community, there is an increasing need to understand the firm level impact of the implementation of these methods. To build the most effective business case for organizations in and out of the software development industry, it is imperative that a case be made to show that the implementation of Agile frameworks has constituted a competitive advantage. This study investigated the organization level performance impact of switching from traditional methods to the use of Agile frameworks. The results showed that changing from a traditional methodology to an Agile framework resulted in higher return on assets and lower operating expense ratios. The interaction between time and methodology for OER, ROA, or revenues in Table 6 did not show a significant difference, indicating that the null hypothesis cannot be rejected. Thus, we cannot say whether performance differs as a function of type of agile methodology. That said, the non-parametric sign test shows that the median improvement in Operating Expense Ratios were highest for Scrum while SAFe seemed to show a slightly higher improvement in Return on Assets. On the whole, Scrum seems to outperform SAFe in terms of operating efficiency (as measured by OER) but lags in terms of ROA.

Table of Contents

Acknowledgements ..... **iv**

Abstract of Praxis ..... **v**

List of Figures ..... **ix**

List of Tables ..... **ix**

Chapter 1: Introduction **1**

1.1. Background 1

1.2. Statement of the Problem 2

1.3. Research Objectives 3

1.4. Research Questions and Hypotheses 6

1.5. Scope of Study 7

Chapter 2: Literature Review **8**

2.1 Introduction 8

2.2 Agile Methods 8

2.3 Origins and formalization of Agile 9

2.4 The Agile Manifesto 10

2.5 Traditional Methods 12

2.6 Agile Methods 15

2.7 Firm level performance 36

2.8 Statistical Methods 46

Chapter 3: Methodology 534

3.1 Experimental Design 534

3.2 Measures 545

3.3 Sample and Data Collection 60

3.4 Study Design 61

Chapter 4: Results 62

4.1 Introduction 64

4.2 Descriptive Statistics 64

4.3 Preliminary Screening Procedures 65

4.4 Primary Statistical Analyses 71

Chapter 5: Discussion of Conclusions 85

5.1 Conclusions 85

5.2 Discussion 85

5.3 Contribution to the Body of Knowledge 88

5.4 Future Research 90

References 91

Appendix I. Data summary. 114

Rejected companies 119

List of Figures

Figure 2-1 Sample Waterfall Project view using a Gantt Chart. 13

Figure 2-2. Model of the PMBOK Process Areas. 14

Figure 2-3 Sample Product Backlog and relative sizes in terms of story points. 23

Figure 2-4 Sample Scrum Board 23

Figure 2-5 Sample Sprint Backlog. 25

Figure 2-6 Sprint Burn Down Chart 25

Figure 2-7 Release Burn Up 36

Figure 2-8 Sample Kanban Board 39

Figure 2-9 SAFe Core Values and Principals 41

Figure 2-10 SAFe Big Picture 43

Figure 4-1 Main effects plots for ROA, OER, and Revenue 78

Figure 4-2 Change Point Analysis graphical results for OER 80

Figure 4-3 Change Point Analysis graphical results for ROA 80

Figure 4-4 Change Point Analysis graphical results for Revenues 81

Figure 4-5 Main effects plots for OER by method 84

Figure 4-6 Main effects plots for Revenues by method 84

Figure 4-7 Main effects plots for ROA by method 84

List of Tables

Table 3-1 Summary of dependent and independent variables 59

Table 4-1 Summary Data 64

Table 4-2 Summary results of Paired T tests 72

Table 4-3 Exact Sign test summary 72

Table 4-4 Friedman's test 73

Table 4-5 Repeated Measures ANOVA 74

Table 4-6 Complex Contrasts for OER, ROA, and Revenues 75

Table 4-7 Post Hoc pairwise comparisons for OER, ROA, and Revenues 77

Table 4-8 Summary Change Point Analysis for OER, ROA, and Revenues 79

Table 4-9 Chow test data 82

Table 4-10 Sign test results by method 83

Chapter 1: Introduction

1.1 Background

Since Agile development methodologies were formalized in 2001, their adoption has spread throughout the software development industry and even begun to be utilized in other industries. The promise of reduced time to market, increased speed, reduction of overhead, adaptability, and improved alignment with customer and organizational needs are widely believed to constitute a significant competitive advantage over firms not utilizing these methodologies.

Increasingly, Agile methods are being adopted outside the software development world. Scrum and other agile methods are becoming popular in marketing and education, and they are expanding throughout the business world (Linders, 2013; Accardi-Petersen, 2011; Hannon, 2014; May, 2016). In education, Bluepoint Education has students using scrum to accomplish their curriculum goals, while organizations like eduScrum have implemented Scrum in secondary and professional training environments (Linders, 2013). Labratoria uses Scrum as well, allowing for 2-3 week sprints in the classroom and allowing for frequent retrospectives and shortening the long feedback loops endemic to traditional education; the Agile Classroom has become their educational model (Prieto, 2016). Walmart is currently transitioning all HR functions to Scrum, following several other organizations including CH Robinson and Verscend (Prieto, 2016; Hoegstron, 2017). Hubspot, Novell, and Pace Communications all use Agile methods for their marketing teams (Ewell, 2011).

As such, if these approaches can be shown as clearly impactful at the organizational level within the Software Development and IT Industry, there are wide ranging implications for organizations outside of that industry. Since the primary driver for publicly owned organizations is delivering and increasing stockholder value, adopting the use of Agile methodologies would thus constitute a core piece of an organization's competitive strategy.

1.2 Statement of the Problem

Agile has become big business. In 2017, Consulting giant Accenture purchased SolutionsIQ with the intention of building its Agile transformation and coaching portfolio (Soh, 2017). Startups with a focus on Agile transformation and agile coaching have proliferated, with small companies like LeanDog showing consistent 3 year revenue growth of over 80% (Inc., 2015). Even small cap organizations are willing to spend millions on transformation services. For example, Verisk Analytics spent over $4 million on external consulting resources in their rollout of the Scaled Agile Framework in 2014 (Neumarker, 2017).

Yet, despite the investment in Agile methods, there is remarkably little data showing empirical impacts at the firm level. While there are dozens of studies extolling the value of implementation of Agile methods, this research has focused at intermediate levels, focusing on projects or functional areas, and even then there is little empirical data (Rico D.,2008).

From the theory of constraints and systems thinking, we also know that optimization at the local level can lead to increased suboptimization at the system level (Trojanowska, 2017; Verma, 1997). As noted by Forte:

> "This is the important point about local optima in complex systems that many miss: local optima are not just suboptimal, as in "not as good as they could be." When combined in an interdependent system, local optima actually make things worse (Forte, 2016)."

This being the case, there is always concern that improvements at the project or functional area may not translate to top level performance, and indeed improvements in one area may negatively impact the overall performance of the organization. As such, measurement at the organizational level is critical to gauge the overall impact to the system (Lakshmi Tulasi, 2005).

Additionally, there have been periodic movements that are viewed by many as 'fads': Methods like Total Quality Management, ISO, Six Sigma, and CMMI have achieved widespread usage, though their impact has often been questioned (Miller, 2002). Despite its use in most organizations, in 57% of organizations, waterfall is still the dominant methodology (Version One, 2016).

1.3 Research Objectives

There is no more important or critical item to organizations than their overall performance and ability to deliver shareholder value. The identification of factors that impact firm level success is of obvious importance to every organization in every industry. A survey of highly regarded journals that publish empirical research on organization showed that over a three year period, 28% of their articles dealt with firm level financial performance as a dependent variable, indicating that this is one of the most critical themes in any form of management research (March, 1997).

Providing hard evidence of economic benefits from the use of Agile frameworks has clear implications for executives and managers. Research going back over 100 years attempts to identify operational frameworks and methods that provide competitive

advantage, with numerous authors and researchers making claims regarding 'what works' (Taylor F. , 1911; Joyce, 2003). Yet, despite extensive research into operating methods and financial performance, the body of scholarly research shows mixed results (Duarte, 2011).

That said, in a survey of 157 companies, it was found that only 23% attempted to identify a causal relationship between non-financial factors (examples include employee turnover, customer satisfaction, and customer loyalty) and firm financial performance, but those same organizations showed on average a 2.9% higher Return on Assets (ROA) and 5.14% higher Return on Equity (ROE) than companies that didn't (Ittner, 2003). In other words, identification and optimization of critical non-financial factors has shown significant impacts to financial performance at the organizational level. The use of Agile methodologies is an example of non-financial factors that could impact firm performance.

Agile methods were developed using the synthesis of multiple fields and using systems engineering methods. Scrum specifically is based in complex adaptive systems theory and leverages over 50 years of best practices. They were developed initially in software development organizations but intended to be industry agnostic (Schwaber, The Scrum Development Process, 1997). Due to the widespread implementation, the creation of Agile methods, and Scrum in particular, is one of the most impactful developments in systems engineering and engineering management in the last several decades (Dyba, 2009).

As such, it is clear that research in this vein has shown significant value, despite the challenges in showing causal relationship at the firm level, as even a small competitive

advantage can be the difference between a thriving organization and one that struggles or fails.

This study is the only one to empirically link directly between the implementation of Agile frameworks and firm performance. Without empirical research assessing bottom line performance, there is little evidence supporting the substantial investment in transitioning existing organizations to Agile methodologies.

This study also provides an analysis on the impact of Agile frameworks as a non-financial factor to complement previous studies on customer satisfaction, quality, and similar measures (Kaplan, 1992; Przasnyski, 2002; Fornell, 2006). Additionally, this study evaluates the implementations of specific frameworks and compares performance of these implementations against each other.

This study seeks to answer the question as to whether the use of Agile has been a competitive advantage to organizations, if a majority of technical firms come to rely on Agile methods for the majority of their projects and teams, it may be that not using Agile methodologies instead puts organizations at a competitive disadvantage. In other words, use of Agile methodologies in the software development and IT industry may be a virtual prerequisite to competing in that space. Because the use of Agile frameworks outside of software development and information technology.

The other critical contribution is the comparison of outcomes between multiple Agile Frameworks. To date, there are no large-scale empirical studies that compare outcomes between competing Agile frameworks.

Agile methods may have grown out of the software development industry, but are widely applicable in other parts of the organization as well. As such, stakeholders

outside of the traditional software development and IT portions of organizations also become key beneficiaries. Showing impact beyond the operational unit makes a strong case to stakeholders in other areas of the Enterprise for the use of Agile methodologies where they are applicable.

Other stakeholders that should be mentioned are management scientists, who need empirical studies showing the linkages between practices and outcomes. While there is a significant amount of data at the project level, empirical data for formalized Agile methods beyond the project level is virtually nonexistent (Kautz, 2014; Rico, 2009). Lastly, use of longitudinal studies is rare in engineering management research as well as in operations research. Novel use of methodologies from other disciplines could help answer many causal related questions in engineering management. Use of statistical tools like Change Point Analysis could highlight techniques that are as yet relatively unknown outside of statistical process control. Firm level performance depends on a large variety of factors, some within the control of the organization and some not. As such, any factors that are strongly tied to improved performance at the firm level are of critical interest to any organization. This study will focus on Operations Expense Ratio (OER), Return on Assets (ROA), and Revenues as the firm level measures.

1.4 Research Questions and Hypotheses

In order to show whether implementation of Agile frameworks constitutes a competitive advantage, we ask the question: Does the implementation of Agile Methodologies lead to improvement in overall firm performance?

In order to evaluate whether the implementation of Agile methodologies constitute a competitive advantage, the following research hypotheses were constructed:

1) OER was lower (improved) for organizations after they changed from a traditional to an agile methodology.
2) The type of framework utilized impacted the degree to which OER improved.
3) Revenues were higher for organizations after they changed from a traditional to an agile methodology.
4) The type of framework utilized impacted the level to which Revenues improved.
5) ROA was higher (improved) for organizations after they changed from a traditional to an agile methodology.
6) The type of framework utilized impacted the level to which ROA improved.

1.5 Scope and Limitations of Study

This study will focus on evaluating historical data to show whether the implementation of Agile Frameworks has translated into improvement in firm level performance. This study is the first to attempt to quantify the actual impact of Agile implementations as they currently exist by using empirical data. It is also the first to provide a large-scale comparison of empirical results for multiple Agile frameworks.

That said, this study does nothing to evaluate how closely any of the organizations embraced Agile values and practices, only whether there was improvement after an Agile transformation occurred.

Chapter 2: Literature Review

2.1 Introduction

This chapter defines Agile development, discusses its history, and provides a detailed description of the most widely used Agile frameworks. Additionally, statistical methods used in the study are discussed as well as operational research that attempts to measure firm financial performance.

2.2 Agile Methods Overview

Though some question remains as to the accuracy of the largest industrial surveys on agile methodologies (Stavru, 2014), Agile methods are now used in the majority of technical organizations. For example, according to the 10th Annual State of Agile Report, 95% of respondent firms utilize agile in some part of their organization, with 43% of those firms reporting the majority of their development teams were using Agile methods (Version One, 2016).

While Agile methods can trace significant influences back to Lean Manufacturing their first formal usage emerged in the 1990's with the advent of Scrum (Rico, 2009) and Xtreme Programming. At this point, the technical practices that have been adopted by many of the Agile frameworks began to emerge. In 2001, key practitioners in the growing movement met and wrote the Agile Manifesto which outlined the core of Agile practices as well as a set of 12 guiding principles of the Agile community (Beck K. e., 2001). Over the next couple of decades, Scrum established itself firmly as the most utilized framework for Agile development. Kanban was adapted from Lean Manufacturing to software development, and other frameworks like the Scaled Agile

Framework (SAFe), Disciplined Agile (DAD), and Large Scale Scrum (LeSS) have arisen (Anderson, 2010; Larman, n.d.; Leftingwell D. e., n.d.; Disciplined Agile Consortium, n.d.). Hybrid methods such as Scrumban are also utilized, and even some methods that use hybrid Agile and Stage gating processes (Conforto, 2016).

Use of Agile methods has continued to grow throughout the world. As of 2016, an estimated 43% of development organizations were predominantly using Agile methods and very few organizations did not have any Agile teams (Version One, 2016). Scrum is by far the most popular methodology both at the team level and for scaling, with the Scaled Agile Framework (SAFe) as the second most popular scaling method. Kanban is the second most utilized method overall (Version One, 2016). Other methods like Large Scale Agile and Disciplined Agile Delivery are starting to gain market share but still have few adherents at this time. Note that most of these methods are not mutually exclusive, as SAFe allows the use of both Scrum and/or Kanban teams, but applies additional constraints at the team levels. Many organizations predominantly use scrum teams, but with some shared services or support teams using Kanban (Al-Baik, 2015; Stoica, 2016). For the purposes of this study, the dominant method is identified.

2.3 Origins and formalization of Agile

While Agile methods were conceived of and implemented initially in the software development industry, the roots of Agile methods go much deeper. Most practitioners trace the agile mindset back to LEAN manufacturing and the works of William Edwards Deming, though the Deming Cycle (Plan Do Check Act) was somewhat derivative of the work of Shewhart of Bell labs, who taught an iterative and incremental approach to improvement (Rigby, 2016). Just In Time and LEAN methods were described explicitly

as early as the 1920's, and iterative models were used through the 70's and 80's, but until the late nineties, process heavy methods, especially the Waterfall model, predominated (Varhol; Ford, 1922).

Through the 90's, the first truly agile approaches took shape. These were lightweight approaches that attempted to allow for easy and rapid adaptation to changing requirements and environments; some of these approaches were Scrum, Xtreme Programming, Crystal Methods, Adaptive Software Development (ASD), Feature Driven Development (FDD), and Dynamic Systems Development Method (DSDM) (Varhol). Thought leaders and practitioners of these methods were the primary participants during the drafting of the Agile Manifesto which formalized the definition of Agile Development. While all of these early methods are practiced to some degree today, only Scrum remains a dominant methodology, though specific technical practices of many of the above frameworks have been adopted as best practices into Scrum and other frameworks (Leftingwell D. e., n.d.).

2.4 The Agile Manifesto

The formalization of what it means to be 'Agile' occurred in Feb. 2001 at the Snow Bird Lodge in the Wasatch mountains of Utah, where a large group of proponents of the increasingly popular 'lightweight' software development methodologies met to attempt to find common ground (Highsmith, 2001). The result was a statement regarding the core of what it means to be Agile. Additionally, there was a list of the guiding principles upon which the statement was made. The Manifesto is as follows:

> "We are uncovering better ways of developing software by doing it and helping others do it. Through this work we have come to value:
>
> **Individuals and interactions** over processes and tools

Working Software over comprehensive documentation

Customer Collaboration over contract negotiation

Responding to change over following a plan

That is, while there is value in the items on the right, we value the items on the left more."

The principles behind the Agile Manifesto are as follows (Beck K. B., 2001):

- "Our highest priority is to satisfy the customer through early and continuous delivery of valuable software."
- "Welcome changing requirements, even late in development. Agile processes harness change for the customer's competitive advantage."
- "Deliver working software frequently, from a couple of weeks to a couple of months, with a preference to the shorter timescale."
- "Business people and developers must work together daily throughout the project."
- "Build projects around motivated individuals. Give them the environment and support they need, and trust them to get the job done."
- "The most efficient and effective method of conveying information to and within a development team is face-to-face conversation."
- "Working software is the primary measure of progress."
- "Agile processes promote sustainable development. The sponsors, developers, and users should be able to maintain a constant pace indefinitely."
- "Continuous attention to technical excellence and good design enhances agility."
- "Simplicity—the art of maximizing the amount of work not done—is essential."
- "The best architectures, requirements, and designs emerge from self-organizing teams."
- "At regular intervals, the team reflects on how to become more effective, then tunes and adjusts its behavior accordingly."

2.5 Traditional Methods

The item with the biggest impact on the way work is structured and tracked is likely the invention of the Gantt chart. Created by process consultant William Gantt some time in 1917, Gantt charts first became widely used as a project management tool to help manage the vastly increased production of munitions during World War I and attempted to reconcile "performance and promises" (Clark, 1922). The Gantt chart provided visualization of the sequencing of efforts in a project and provided a planning and tracking tool that became ubiquitous, spreading throughout the military before the end of the war (Black, 2014). Over the next several decades, Gantt charts were the primary tracking mechanism used in the construction of the Hoover Dam, and the Interstate Highway system: even where Gantt charts were not used, the systemization of management practices and development of management science in the early part of the 20th century firmly left a sequential, heavily planning focused impact on the way work was done throughout the world (KIDASA Software, n.d.).

Two additional events occurred to have a tremendous impact on the organization of work. First, in 1970, William Royce coined the term 'Waterfall' as a software development process. Despite noting in the same paper that in his experience that simple waterfall model didn't work well on large projects, its simplicity appealed to many managers and it was quickly widely adopted as the primary method of software development (Kessel, 2013; Royce, 1970). The waterfall model matched the way many non-software projects were managed and was popular in part because everything flows logically from the beginning of a project through the end. Increased computing power

also allowed for easy creation of more complex Gantt charts to track progress as well. The waterfall model is characterized by significant up front planning, heavy documentation, and often incurred significant lag time between the creation of a defect or issue and attempted resolution. It was important to identify all requirements up front and did not readily allow changes once design and implementation were underway. Phase gate processes were often introduced and incurred significant slack in the system (Sutherland, 2014; Hughey, 2009). Figure 2-1 shows an example of a typical Gantt chart generated from Excel.

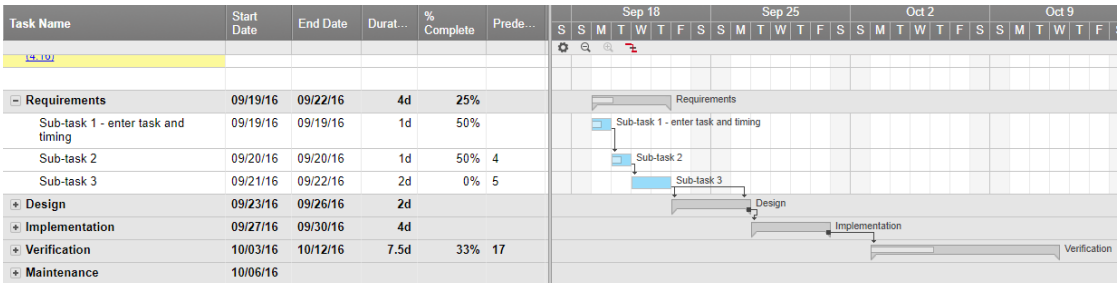


Figure 2-1 Sample Waterfall Project view using a Gantt Chart.

The other incident of note was the creation of the Project Management Institute (PMI) in 1969. Most known as a certification body for project and program managers, PMI compiled a comprehensive body of knowledge for project management processes, principles, and best practices (Sliger, 2008). It should be noted that the advent of the primary Project Management organizations like PMI and its analog in Europe, INTERNET which was the progenitor of International Project Management Association and Association of Project Management (IPMA and APM respectively) were primarily founded by and focused on project scheduling and controls in the early years (Weaver, 2007).

The Project Management Body of Knowledge (PMBOK) and its most popular certification, the Project Management Professional (PMP) rapidly became the de facto project management resource within the US. The PMP became a required certification for project managers in the federal government and other industries. As such, the processes and methods espoused in the PMBOK became utilized by project managers in almost every industry and became the de-facto standard.

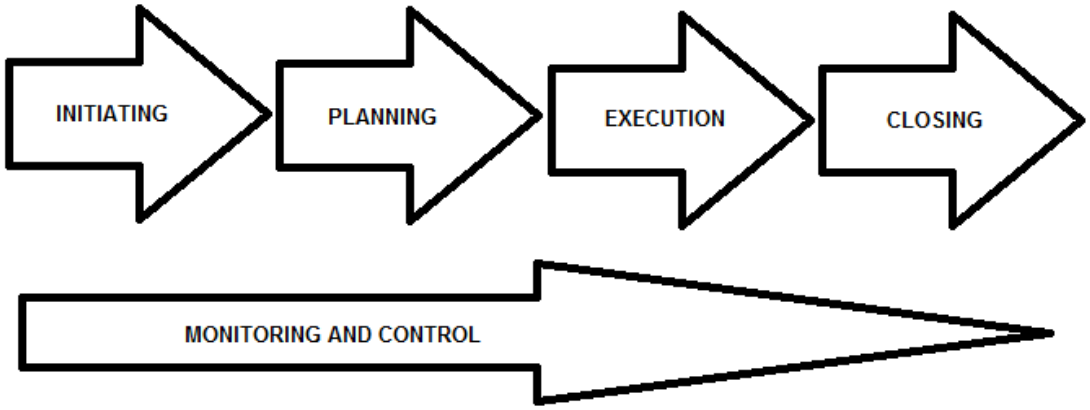


Figure 2-2. Model of the Project Management Body of Knowledge Process Areas. Based on the Project Management Body of Knowledge (PMBOK).

While it should be noted that the PMBOK does not explicitly advocate the 'waterfall methodology', but rather identifies 47 processes that exist in most (but not all) projects within 5 primary process areas, the nature of those process areas and the prevalence of the waterfall approach lead to almost universal adoption of a waterfall approach from PMI members and throughout the federal government, its contractors, and the vast majority of project management through the late 2000's (Walenta, 2015; Sliger, 2008).

PMI proponents admit that earlier versions of the PMBOK make it difficult to see that Agile methods are supported, but that from the 2004 version on, there have been attempts to make the PMBOK more open to Agile methods, supported by multiple articles in PMI's PM Network magazine starting in 2005. In 2012 PMI began offering their own Agile certification as well (Sliger, 2008). That said, the PMBOK is still heavily perceived as a waterfall approach to this day (Walenta, 2015).

Many Agile thought leaders still consider the PMBOK guidance to be, in large part counter to the Agile principles. As the PMBOK offers significant guidance regarding extensive documentation and has generally been interpreted to require extensive up front planning, including the development of not only a detailed Work Breakdown Structure (WBS) and Project Management plan, but also recommending the development of a detailed Communications Plan, Quality Plan, attachment of cost estimate data to the WBS, Requirements traceability matrix, Project Charter, Stakeholder Management strategy and plan, detailed scheduling and cost development, and Risk Management plan (Association of Modern Technology Professionals, n.d.; Fernandez, 2009). .

2.6 Agile Methods

2.6.1 Scrum

The first Scrum team was formed at the Easel Corporation in 1993 and the process was iterated and refined over the next several years before being presented at an Association for Computing Machinery research conference by Jeff Southerland and Ken Schwaber in 1995 (Sutherland, 2014). Scrum drew primary inspiration from two sources. First, a groundbreaking article from Tacheuchi and Nonaka (1986) described the traits of

the most effective project teams through a significant meta-analysis. They found that the most effective teams had a strong shared Vision, were cross-functional, and had a high degree of autonomy. They also described teams operating in lockstep and were the first to use the rugby analogy that Sutherland and Schwaber would eventually adopt when naming Scrum (Takeuchi, 1986).

The other is an anecdote told by Jeff Sutherland where Rodney Brooks, a professor of Artificial Intelligence at MIT explained how despite spending billions of dollars and many years trying to build bigger, more powerful computers with huge databases, artificial intelligence (AI) wasn't progressing effectively, but his new robots had a built in brain for each of the six limbs, and a central processor had a few simple rules. The central processing chip knew the rules and would provide feedback to the individual brains. Each time the machine was turned on, it learned to walk for the first time. In other words, the individual legs acted as autonomous agents and quickly learned to collaborate and move efficiently and effectively. Sutherland purportedly asked "What would happen if we could come up with a simple instruction set for teams of people to work together just like those legs. They would self-organize and self-optimize, just like that robot" (Sutherland, 2014).

Essentially, Sutherland came from a biostatistics background where part of his dissertation was regarding biological systems as complex adaptive systems. As he moved into academia and later into the corporate world, he pulled from research in all areas, starting with complexity theory, but also looking at all the studies in psychology, motivation, knowledge worker productivity, team dynamics, multitasking, Lean manufacturing, American Special Operations Forces, leadership, system dynamics and

system thinking, experience and training from his time as an Air Force fighter pilot, and the quality system management works of William Edwards Deming. While studies regarding how people work most effectively were a rich research topic going back to World War 2, nobody had synthesized and combined the research effectively (Sutherland, 2014).

Although Scrum predates the Agile Manifesto, first and foremost it adheres to the guidance in the manifesto and the principles behind it. Scrum is a team level empirical process that allows each team great flexibility in how they operate and deliver. Multiple teams exist in a rapidly changing environment and allowed maximum flexibility, as evolution favors those with maximum exposure to environmental change and deselects those who are insulated from the environment (Schwaber, 1997).

2.6.1.1 Overview of Scrum

Work is organized in short cycles called sprints that are from 1 to 4 weeks in length, though most teams tend to utilize 2 or 3 week sprints. Before the sprint starts, teams estimate how much they can do in the time frame and pull the work into the sprint based on priority. In this manner, they limit their work in process (WIP). During this work cycle, management does not interrupt. The team is self-reporting and impediments are systematically removed. At the end of each sprint, the team reflects on its performance and builds in an inspect and adapt effort to continuously improve and adapt (Deemer, Scrum Primer 2.0 A lightweight guide to the theory and practice of scrum, 2012).

2.6.1.2 Scrum Roles

As initially proposed, Scrum consists of three roles, three ceremonies, and three artifacts.

- Roles: Product Owner, Scrum Master, and Team
- Ceremonies: Sprint Planning, Sprint Review or Demo, Sprint Retrospective
- Artifacts: Product Backlog, Sprint Backlog, Burn Down Charts

Later, as Scrum was used at larger scales, the Release Burn up Chart was added as an additional artifact (Ebert, 2017).

The team and its dynamics are the cornerstone of Agile delivery. While top performing individuals can be as much as ten times as efficient as good employees while maintaining the same quality of work, the best performing teams can be over 2,000 times as fast, again with the same quality of work (Sutherland, 2014).

The Team is cross-functional and consists of 5-9 members. Ideally, there are no titles, though in practice that is rarely the case (Deemer, 2012). The team and team dynamic are absolutely core to scrum, and teams should be self-managing and self-organizing. Teams should be co-located and team members should be dedicated, not splitting their time between multiple teams, as dividing focus between teams also makes it more difficult to control priority and limits the self-organization of the team (Schwaber, The Scrum Development Process, 1997). In the case of larger projects, multiple teams can work on the same project instead of increasing the size of the team to 10 or more members. Studies continually show greater productivity and communication with smaller team sizes and over 10 members shows a significant degradation in team performance (Armel, 2012). The team is an autonomous unit that has a high level of control on how the work is performed. They also have control over how much work is pulled into a given sprint (Rubin, 2013).

The Product Owner serves as the voice of the business and is responsible for providing return on investment for the work done by the team. He or she must identify product features and prioritize work in preparation for the next sprint, while providing guidance to the team regarding intent of existing work. In a commercial environment, they may have profit and loss responsibility for a product line, though at times a customer will actually serve as the product owner (Deemer, 2012). The Product Owner provides a single point of prioritization for the team, allowing them to minimize task and context switching as well as limiting multitasking and other distractions to the team (Bennett, 2014).

The Scrum Master is a servant leader to the team. Their role is to do whatever is in their power to help the team, product owner, and organization to be successful. As such, they are responsible for removing impediments of all types, protecting the team from external interference, and helps coach the team, Product Owner, and other stakeholders on effective use of scrum. It is highly recommended that the Scrum Master be a full time, dedicated role (Deemer, 2012).

2.6.1.3 Scrum Ceremonies

There are multiple ceremonies in Scrum as well. Before the Sprint, there is a Sprint Planning session, and at the end there is a Sprint Review and a Retrospective. There is also a daily ‘Scrum’ or ‘Standup’ meeting and often larger organizations use a Scrum of Scrums to aid in communication and coordination (Deemer, Scrum Primer 2.0 A lightweight guide to the theory and practice of scrum, 2012).

Sprint Planning occurs at the beginning of every Sprint. The Product owner and the team agree on the goal of the sprint and the team pulls work into the sprint based on

priority, pulling in enough work that they believe they can realistically complete the work and maintain at a sustainable pace. The team may need to estimate the size of the work items, often framed in a User Story format. Each work item is broken down into individual tasks with discussion regarding architecture and implementation. Usually, the tasks are sized in hours (Rubin, 2013).

The Daily Scrum or Standup is a short duration meeting held every day that includes the Scrum Master, all Team Members, usually the Product Owner, and any stakeholders that need to be there. The meeting should not take more than 15 minutes and is considered an inspect-and-adapt activity, usually consisting of three questions answered by each team member, often followed by in depth discussion on one or more topics (Deemer, Scrum Primer 2.0 A lightweight guide to the theory and practice of scrum, 2012; Rubin, 2013). The questions are:

1. What have I accomplished since our last meeting?
2. What will I work on next?
3. Are there any impediments preventing me from getting the work done?

It needs to be made clear that this is not to be a status meeting, but rather a point for coordination and assessment regarding progress towards the sprint goal or goals. It is also not a venue for deep problem solving, but should highlight problems that can be addressed in a follow-on meeting (Rubin, 2013).

The Sprint Review is an opportunity to evaluate the product being built. This usually includes a demo, and can have any number of stakeholders present in addition to the Product Owner, Team, and Scrum Master. This is often an opportunity for customers and other stakeholder to provide real time feedback directly to the team to allow for better

product development in the future, as well as for the team to provide insight on development decisions (Rubin, 2013).

While the Sprint Review offers an inspect-and-adapt venue for the product, the Retrospective allows the team to adapt the process and to continuously improve. The Product Owner, Scrum Master, and Team come together to evaluate what the team is doing well, what impediments are present, and evaluate new methods or approaches to continually improve. With relatively short sprint duration, the team has many opportunities to adapt and improve over time (Rubin, 2013; Schwaber, 1997).

The Scrum of Scrums is widely considered a possible method of scaling Scrum, and is often used as a way of coordinating between multiple scrum teams, especially in situations where there are dependencies. It is analagous to the daily Scrum, but with representatives from multiple teams in attendance (Agile Alliance, n.d.).

While not always formally presented as a ceremony in Scrum, it is recommended by many coaches and Agile leaders to have periodic Backlog Refinement or Grooming sessions, where the team will sit with the Product Owner and discuss work that will be slated to go in future sprints. They offer an opportunity for the team to ask questions and obtain clarification regarding future work from the Product Owner, as well as a venue for collaborative definition of work to be completed in future sprints (Deemer, 2012).

2.6.1.4 Scrum Artifacts

There are several artifacts used in Scrum for the managing and tracking of work. First, there is the concept of a Product Backlog, or the overall list of things that the Product Owner has that need to be done. This is structured in a stacked priority order that makes it easy for teams to pull the most important items in order when conducting Sprint

Planning. These items are estimated in terms of size, generally using relative sizing techniques and quantified in terms of story points. Figure 2-3 shows a sample Product Backlog with a prioritized list of User Stories and estimates in terms of Story Points. Story points are numerical values without units that represent the relative size of efforts captured on the Product Backlog and generally expressed using a modified Fibonacci sequence, with the smallest effort being 1, followed by 2, 3, 5, 8, 13, and 21 respectively (Coelho, 2012). The relative sizes of the User Stories are used as an input to Sprint Planning, and the amount of Story Points completed per Sprint is the Team's Velocity, a measure of productivity to measure the team's throughput and continuous improvement (Pomar, 2014).

During Sprint Planning, the team builds a Sprint Backlog, a list of stories, items, or features pulled into the Sprint, and also a list of tasks associated with those items. This backlog is placed visibly on a Scrum Board so that task progress can be visually tracked and communicated (Rubin, 2013).

In the sample Scrum Board pictured in Figure 2-4, the items under Story are units of work expressed in terms of value. The items in the other columns represent tasks within the user stories and are color coded accordingly. Tasks are pulled to In Progress when work is begun, and to Done when a task is completed.

Backlog item	Estimate
As a casino owner, I want to have single machine progressives so that…	5
…single bank progressives…	8
…full casino progressives…	5
As a casino employee, I can run revenue reports by machine….	3
Improve exception handling	8
…	
…	

Figure 2-3 Sample Product Backlog and relative sizes in terms of story points.

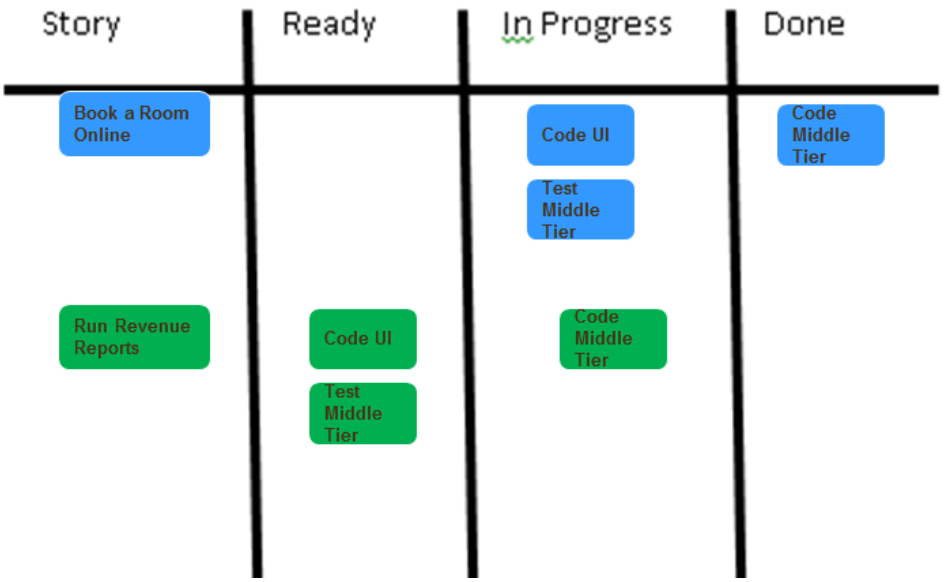


Figure 2-4 Sample Scrum Board

During Sprint Planning, Task durations are estimated in hours, establishing an overall estimate for the work to be done within the Sprint. This is captured in the Sprint Backlog.

Figure 2-5 shows a sample Sprint Backlog. In the Sprint Backlog, the value for each task represents time remaining to complete the task (in hours). In general, sunk costs are not tracked (time spent), but time remaining is tracked and recorded daily (Cervone H. , 2011). Time remaining can increase if a task later proves to be larger than the initial estimate (as in Implement Long Poll 2C, where the time remaining increased from 8 hours to 16 hours between Monday and Tuesday).

The remaining work to be done on tasks that are in progress is re-estimated on a daily basis, new tasks are added, and irrelevant tasks are removed. When they are re-estimated, they are to show only the hours remaining, regardless of effort spent. The total hours remaining are reflected on the Sprint Burn Down chart, as shown in Figure 2-6. Note that addition of tasks or tasks that are re-estimated to be higher than the previous time remaining can actually cause the burn down to go up from one day to the next (Deemer, 2010).

Tasks	Mon	Tues	Wed	Thur	Fri
Implement long poll 1B	8	4	8	0	0
Update WPF page	16	12	10	7	0
Implement Long Poll 2C	8	16	16	11	8
Write online help	12	0	0	0	0

Figure 2-5 Sample Sprint Backlog

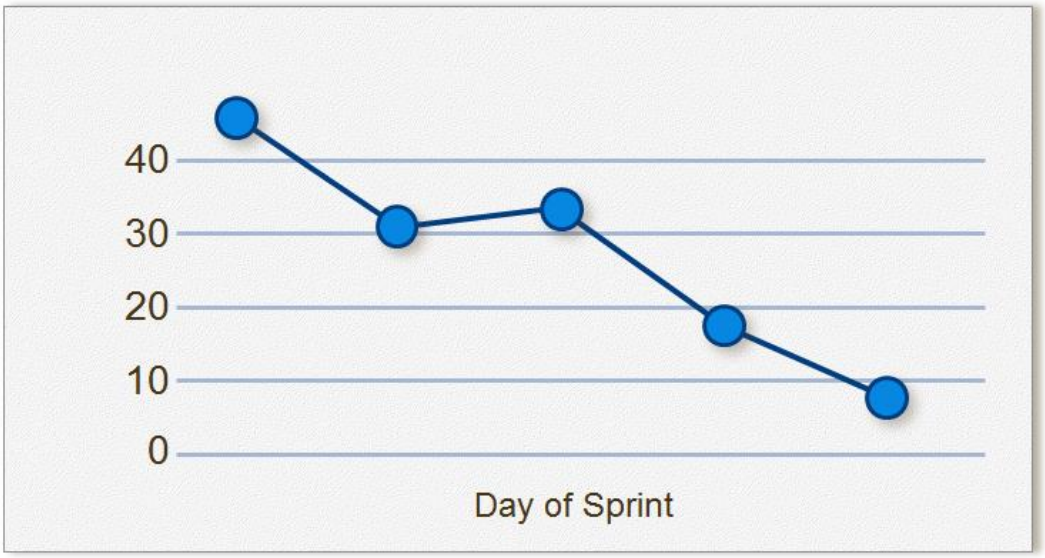


Figure 2-6 Spring Burn Down chart

Figure 2-6 shows a graphical representation of the time remaining from figure 2-5. Note that re-estimation of time remaining can result in an increase in total hours remaining. This provides a graphical representation for the team to track performance against their initial sprint plan. This is called a Burn Down chart because at the beginning of each Sprint, the total estimated hours to finish the work allocated to the sprint is at its peak. Over the course of the Sprint, as hours are re-estimated, the total number of hours remaining in the Sprint will drop, ideally hitting zero hours remaining at the end of the Sprint, indicating that all planned work is completed. This allows external stakeholders to monitor the progress of the Sprint easily (Cervone, 2011).

When tracking very large efforts that cannot be completed in a single Sprint and/or that are split between multiple teams, often a Release Burn Up chart is utilized to track progress. Figure 2-7 is an example of a Burn Up chart generated in Jira. Unlike a Burn Down chart, the Release Burn up represents the total planned scope of a release, epic, theme, or other larger scope of work, usually computed by adding the Story Point sizes

for all User Stories in the effort. In the plot below, this is represented by the light blue line at the top of the plot. Note that as changes are made to the scope, the level of this line can increase or decrease. This line represents the total scope in order to be considered done with the effort. The x axis is a listing of the next several sprints. At the conclusion of each Sprint, the Velocity is recorded, allowing stakeholders to see progress to the overall goal and to predict when the release will be completed based on the slope of the line formed by the Sprint data (dark blue below) and the current scope of the effort. This provides a clear visual representation of progress toward the goal while allowing scope to float as needed (Rubin, 2013; Heredia, 2014).

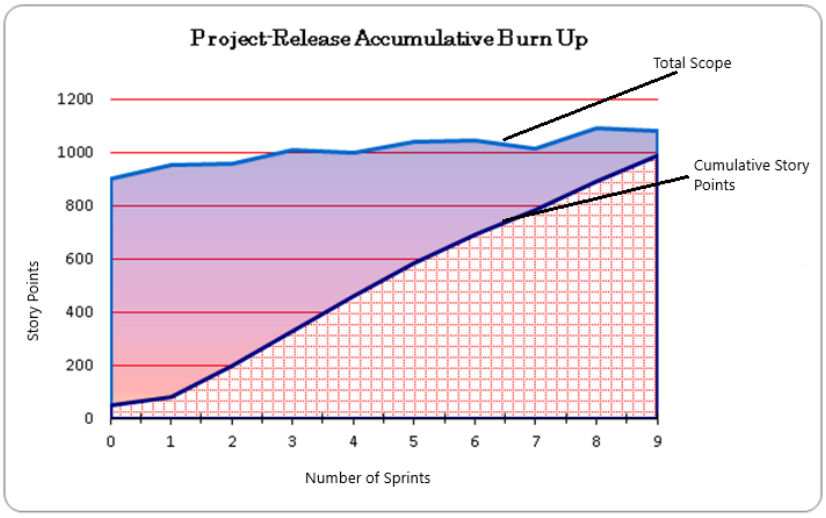


Figure 2.7. Release Burn up chart.

2.6.2 Kanban

Kanban is a LEAN technique that generally meets the principles of the Agile Manifesto. Kanban is a core part of the Toyota Production System (TPS), but was made widely accessible to software development organizations at the end of the 2000's and early 2010's (Anderson, 2010).

Translated from the Japanese, Kanban literally means ‘visual signal’. Kanban is characterized as a ‘pull’ system that manages work in progress (WIP) explicitly and uses queues in order to manage work flows (Thun, 2010). Kanban provides a way to communicate between processes and facilitate the efficient operation of your ‘pull’ system. There are four core principles of Kanban; visualize work, limit work in progress, focus on flow, and continuous improvement (LeanKit, n.d.). In the software development literature, these traditional principles have morphed slightly to five: visualize the workflow, limit WIP, manage flow, make policies explicit, and implement feedback loops (Al-Baik O. M., 2015).

Visualize work means making the process visible and apparent so it becomes easy to identify bottlenecks. Traditionally there are three types of Kanban, all focused on the visualization of the process. The earliest Kanban’s were utilized in inventory control, where an empty space was an indicator for restocking. This is also utilized in many large manufacturing environments, where large carts are queued at each workstation as they move through the process, and a worker pulls the next in line when they are finished what they are working on, provided they are not exceeding the queue for the following step in the process (New, 2007). Similarly, empty containers are also used (Lean Lab, n.d.). For example, at BAE Systems Space Systems Electronics Electromechanical Assembly lab we would prioritize the next several sets of assemblies and kit the parts, placing them in priority order in the staging area. As an operator finished their current work, they would move the completed assembly to the inspection queue. If the inspection queue was full, the assembly could not be moved, and the operator was not to pull another assembly in process. The operator could choose to inspect a previous subassembly in the queue to

make room. This eliminated overproduction at the bottlenecks and kept the flow of the system high.

If, on the other hand, they did not fill the inspection queue, they would pull the next most important assembly to their station and begin manufacture. To state this another way, if the inspection queue was full, they could not pull more work into manufacture until the backlog at inspection was addressed.

In the software development world, generally either a physical board maps out the process steps and cards are used to represent the work involved or an electronic version of a board is utilized to display the workflow. As the work each card represents moves through the process, the card is moved to provide a visual representation to all stakeholders on the status of the WIP (Tanner, 2017).

Limiting WIP is another key concept in Kanban. In LEAN thinking, there is a strong drive to eliminate Muda, or waste from the system. Building long queues is considered a big waste. Overproduction when there is a slower process downstream, a bottleneck, or an impediment makes the overall system less efficient. Lower WIP also reduces multitasking and context switching, and makes sure everything started gets done and doesn't languish in progress for an excessive period of time. The idea is finish what you start, then work on the next most important thing and get it to done. It is better to have one thing completely finished than a dozen halfway completed (Harrison, n.d.; Anderson, 2010; Schaller, 2005). In Kanban systems, WIP is limited explicitly i.e. only a certain number of items are allowed to be in progress at any given time.

In Figure 2-8, note that the number in each column is an explicit WIP limit and applies to the total number of cards in each column. This means that items in the Done

portion of the column count against WIP limits as well, as long queues and waiting time is considered a significant waste in Lean thinking.

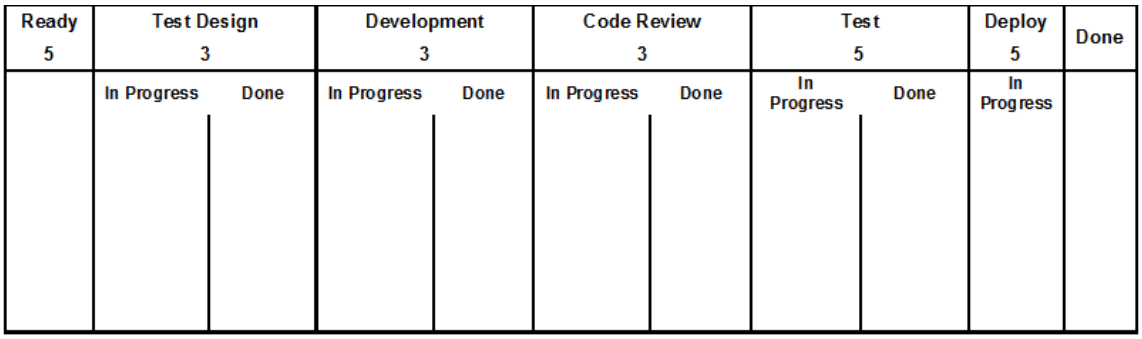


Figure 2-8 Sample Kanban Board

Kanban systems can show bottlenecks very clearly. A focus on flow and addressing bottlenecks and constraints effectively makes the system more efficient. Continually addressing those bottlenecks and improving the process allows the system to adapt effectively over time (LeanKit, n.d.).

As mentioned previously, Kanban is a pull system. Most traditional systems are 'push' systems, in which raw materials or backlog items go through each step in the process, usually with local optimization and often leading to significant overproduction at some steps and building huge queues at others. When large queues of unfinished work are awaiting test, a common issue in software development, the delay in being able to address issues increases the difficulty and cost of repair (Shalloway, 2011). Pull systems only allow work to progress if there is 'room' for it. Scrum manages WIP by pre-planning for a very short duration of work, while Kanban explicitly limits the number of things that can be in a certain process area at a time (Anderson, 2010).

Another key point about Kanban is that it is a continuous flow system. Instead of pre-planning for a sprint, work is prioritized and fed to the system and run through

whatever process the team has implemented. This can be a waterfall process, but is managed in a Lean manner.

General Kanban rules are as follows. A process produces only what a later process needs, and never push production to later processes. The later process informs the earlier what to product, the later process pulls from the earlier process, and defects are not passed through and are addressed immediately. Everything goes on the Kanban board (Lean Manufacturing Tools, 2017).

It should also be noted that a Kanban can be used for virtually any process that is in existence. The first step to implementation is to map out the current process. Then, apply WIP limits. Identify bottlenecks and issues, and continuously improve the process. This allows for continuous, incremental improvements without requiring a complete organizational overhaul (Anderson, 2010).

2.6.3 Scaled Agile Framework (SAFe)

While Scrum, Kanban, and other methods had become widely utilized, the primary discussion on large scale agile implementations was usually limited to talking about a 'scrum of scrums'. In an attempt to provide a scaling solution appropriate for larger organizations, in 2011 Dean Leftingwell rolled out the first formal version of the Scaled Agile Framework (SAFe) (Leftingwell D. , 2017). SAFe is an empirically derived, relatively prescriptive framework that nevertheless recommends adapting to your given organization. The framework is constantly evolving, in some cases addressing criticism and always incorporating new data from empirical implementations (Woodward, 2013). SAFe differs from most of its predecessors in that it applies a framework around existing frameworks, allowing for and implementing the use of both Scrum and Kanban. It is

based on Lean and Agile concepts (Scaled Agile, 2015). In addition to its basis in Lean methodologies and the theory of constraints, it also borrows many technical practices from Xtreme Programming (Scaled Agile, 2015). Figure 2-9 summarizes the key principles of SAFe.

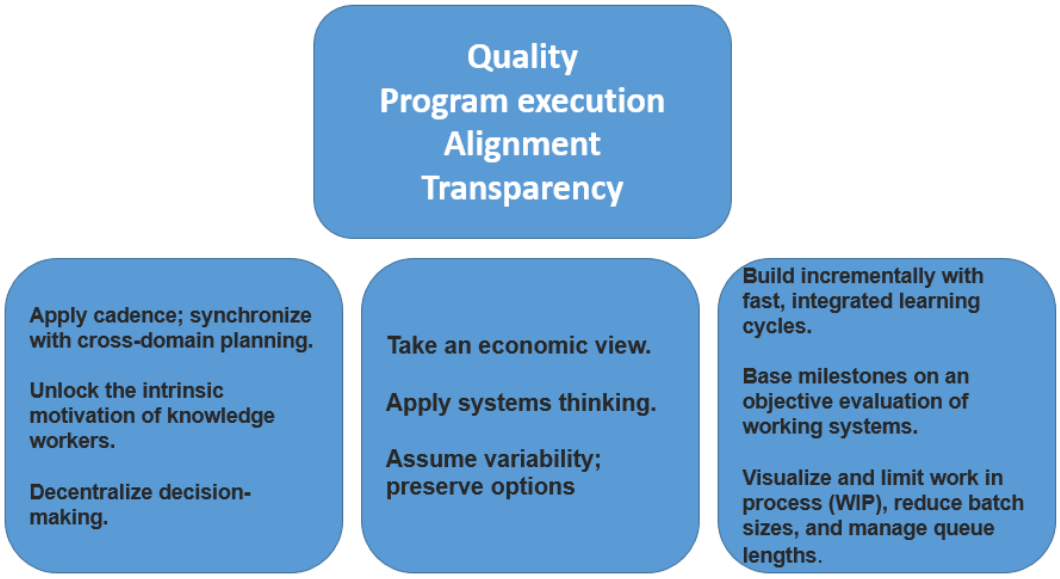


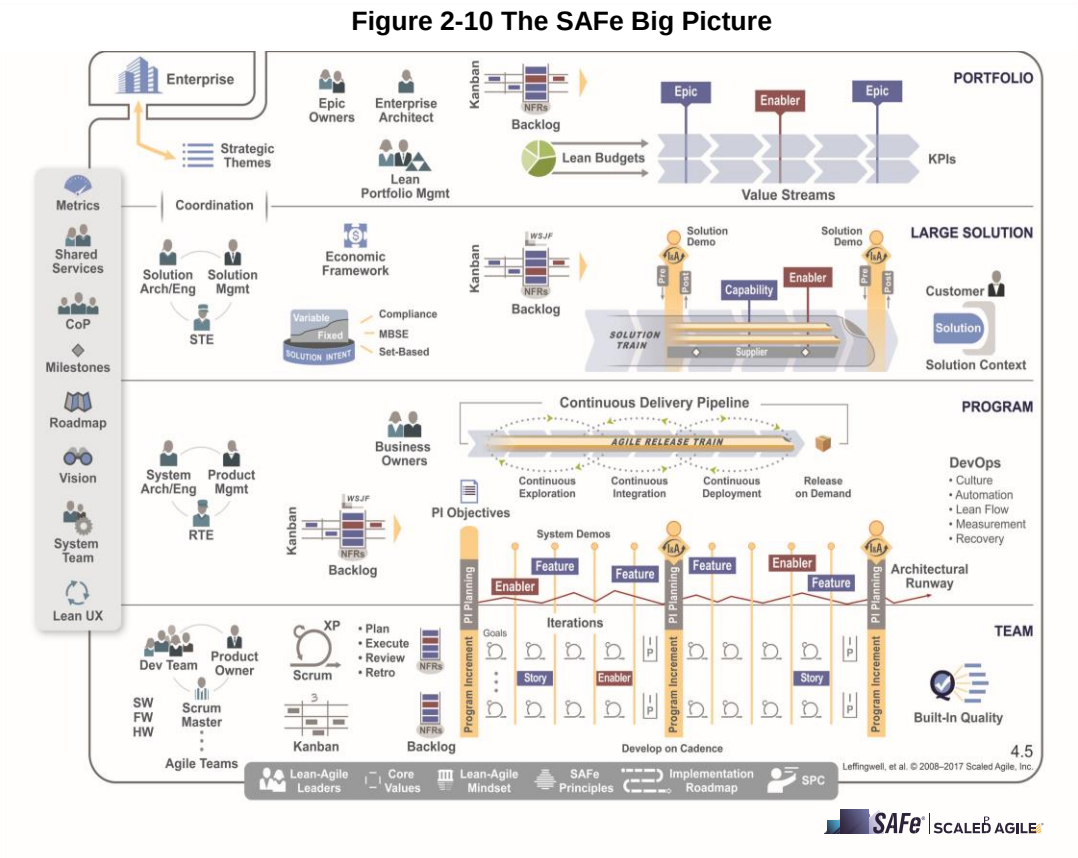
Figure 2-9 SAFe Core Values and Principles

Essentially, SAFe breaks the organization out to 3 or 4 levels, depending on the size of the organization. The lowest level is the team level, usually comprised of Scrum and Kanban teams. Above the team level is the program level, organized around the flow of value within a defined product area (called a value stream) (Turetken, 2016). This is the primary vehicle of delivery. At the program level, SAFe introduces the concept of Release Train, which is essentially a group of several Agile teams. A new role was also created, that of Release Train Engineer, who acts as a higher level Scrum Master, essentially driving the Release Train as a Scrum Master often drives the Sprint. Another new role, the Product Manager focuses on program level prioritization and roadmaps.

Above the program level is the portfolio, usually composed of multiple program teams. A recent addition to the model is Large Solution SAFe. Large Solution SAFe generally has value streams that cannot be fully supported with only one Release Train, as train size is capped at 125 people to keep communication between stakeholders manageable. In Large Solution SAFe, multiple Release Trains operate mostly independently within the same Value Stream. (Hayes, 2017; Scaled Agile, 2017).

These teams are working in a Program Increment (PI), which generally consists of 4-5 sprints. This is followed by an Innovation and Planning (IP) Sprint, in which functional work is generally not planned, but teams are encouraged to work on their own innovations. Within the IP Sprint, the entire Release Train comes together for a 2 day planning session during which they take the prioritized features from the Product Management team and pre plan the sprints in the PI, pulling the work into the sprints in priority order. Performance of the previous PI is also reviewed at the PI Planning session. At the end of the PI Planning session, the teams, product ownership, and senior management agree to the roadmap for the next several sprints. In addition to PI Planning, Inspect and Adapt workshops are recommended to address program level challenges or issues above and beyond the team level retrospectives that occur with each sprint. The SAFe Big Picture, pictured in Figure 2-10 is a graphical depiction of the entire framework (Scaled Agile, 2017).

Figure 2-10 The SAFe Big Picture



SAFe has come under significant criticism from other leaders in the Agile community. Ken Schwaber wrote a scathing article about SAFe and other prominent Agilists have expressed skepticism, concern, or outright disdain for the framework (Sedge, 2014; Schwaber, UnSAFe at any speed, 2013; Adkins, 2014). In many cases, initial criticism was tempered after attending the classes, and the general concerns in implementation of SAFe are that it requires underlying Agile behavior and is likely to be implemented in organizations that have already struggled to implement said behavior. Likewise, because it is more *comfortable* there is concern that most implementations will focus over Processes and Tools over individuals and interactions, in violation of the Agile Manifesto (Adkins, 2014; Sedge, 2014).

2.6.4 Large Scale Agile (LeSS)

Formed by Craig Larman and Bas Vodde in 2005, the Large Scale Scrum (LeSS) framework is an attempt to strike a balance between principles and practices similar to that struck by Scrum. The LeSS framework seeks to be less prescriptive than other scaling methods to provide some rules but with a focus on principles and experimentation (Srinivasasan, 2016).

In addition to drawing from Scrum specifically, LeSS draws from queuing theory, empirical process control, Lean, and Systems thinking (Srinivasasan, 2016). They offer two different frameworks. The first supports up to 8 teams and adds things like multiple team sprint planning, open space, and scrum of scrum meetings, with teams primarily structured as feature teams. There is still only one product owner and one product backlog, and teams coordinate planning, reviews, retrospectives, and grooming sessions. The Product Owner is more of a connector of teams to stakeholders, with focus primarily on prioritization, not clarification. The second framework is for projects that require more than 8 teams and adds product level sprint reviews and retrospectives as well as adding multiple Product Owners (Rabon, 2015). As such, this is widely considered the most 'agile' scaling methodology.

Operationally, teams have a shared product backlog that is pulled into individual teams during iteration planning

2.6.5 Disciplined Agile Development (DAD)

Disciplined Agile, initially Disciplined Agile Delivery (DAD) is a lightweight framework that provides scaling solutions as well as organizational guidance to become more Agile. The idea is to help organizations streamline to support overall agility by

addressing Delivery, Dev Ops, Architecture, Program Management, Finance, and other relevant pieces of the organization. DAD is unique in that it styles itself as a Decision Framework, meaning it tries to capture many experiences, tradeoffs, and offers multiple approaches and explains the empirical results (Woodward, 2013). Additionally, it is considered a hybrid approach in that it draws from many more traditional methods and practices than other agile frameworks (Rabon, 2015). Like SAFe, it also has a lot in common with the Rational Unified Process (RUP), a framework that has essentially disappeared. That said, it has not received the level of criticism that SAFe has, likely due to its lower and slower growing market share (Version One, 2016).

In order to incorporate lessons learned when working at the Enterprise level and address areas outside of software development, within the DAD framework the Agile Manifesto and underlying principles has been rewritten, something a few others have done (Ambler, 2014; Ambler S.). The updated Manifesto and Principles are as follows:

> **"Individuals and interactions** over processes and tools
>
> **Consumable solutions** over comprehensive documentation
>
> **Stakeholder collaboration** over contract negotiation
>
> **Responding to change** over following a plan"

1) "Our highest priority is to satisfy the stakeholder through early and continuous delivery of valuable solutions."
2) "Welcome changing requirements, even late in the solution delivery lifecycle. Agile processes harness change for the customer's competitive advantage."
3) "Deliver consumable solutions frequently, from a couple of weeks to a couple of months, with a preference to the shorter time scale."
4) "Stakeholders and developers must work together daily throughout the project."

5) "Build teams around motivated individuals. Give them the environment and support they need, and trust them to get the job done."

6) "The most efficient and effective method of conveying information to and within a delivery team is face-to-face conversation."

7) "Consumable solutions are the primary measure of progress."

8) "Agile processes promote sustainable delivery. The sponsors, developers, and users should be able to maintain a constant pace indefinitely."

9) "Continuous attention to technical excellence and good design enhances agility."

10) "Simplicity – the art of maximizing the amount of work not done – is essential."

11) "The best architectures, requirements, and designs emerge from self-organizing teams."

12) "At regular intervals, the team reflects on how to become more effective, then tunes and adjusts its behavior accordingly."

13) "Leverage and evolve the assets within your enterprise, collaborating with the people responsible for those assets to do so."

14) "Visualize work to produce a smooth delivery flow and keep work-in-progress (WIP) to a minimum."

15) "Evolve the enterprise to support agile, non-agile, and hybrid teams."

2.7 Firm level performance

2.7.1 What types of factors impact firm level performance

While it can be difficult to isolate factors that impact performance at the firm level, the factors that impact a firm's performance can be broken down into three categories. Those categories are organizational factors, environmental factors, and people factors (Hansen, 1989). Hansen built two models to evaluate how much of the variance in firm Return on Assets (ROA) was impacted by each model individually and also built an integrated model to test independence of the respective models. The models used were

an Economic model and an Organizational model. The Environmental model consisted of the following predictor variables: Industry profitability, market share, and firm size. The Organizational model utilized the following predictor variables: communication flow, emphasis on human resources, decision making practices, organization of work, job design, and goal emphasis. Findings showed that the Economic model and the Organization model acted independently, with little difference between the results of the Integrated model and the individual models. The Organizational model contributed 38% to firm performance, while the Economic model contributed only 19% (Hansen, 1989). The implementation of Agile methods involves a complete overhaul of the organizational factors noted above. One of the primary concerns in this study is the potential impact of the economic factors. That the relative impact to firm financial performance is nearly two to one in favor of the organizational factors strengthens the case for causality due to implementation of Agile methods.

2.7.2 Research in specific factors that impact firm level performance

There is an ever-increasing body of knowledge attempting to identify factors that potentially form a causal relationship with firm financial performance. A survey of well regarded journals that publish empirical research on organizations that included Administrative Science Quarterly, the Academy of Management Journal, and the Strategic Management Journal found that 28% of their articles attempted to establish a causal link from intermediate factors to firm financial performance (March, 1997). Links have been established between external knowledge usage, market orientation, leveraging of information systems, strategic flexibility and firm performance (Bapuji, 2011; Wei, 2014; Zhang, 2005). Others evaluated the impact of innovation management systems,

ERP implementations, RFID implementation, and Automated Teller Machine (ATM) investment as factors impacting firm level financial performance (Chang, 2011; Hwang, 2015; Mir, 2016; Hung C. S., 2012). Wu and Wang evaluated the impact of resource based view (RBV) and the transformation of resources at the firm level (Wu, 2007). Many studies have addressed board composition as a driving factor in firm performance (Duran-Encalada, 2015; Campbell, 2008; Ongore, 2015). One of the most recent and comprehensive studies of this kind found a significant positive relationship between Return on Assets (ROA) and CEO tenure, board independence, ownership concentration, and CEO duality (Rostami, 2016).

The most relevant studies evaluate the implementation of operational frameworks. The Balanced Scorecard was among the first methods that attempted to suggest a causal relationship between customer satisfaction and firm performance, with later studies attempting to tie Total Quality Management (TQM) and other quality measurement tools to firm performance, yet even today, there has been little research into the relationship between quality or customer satisfaction and firm performance (Kaplan, 1992; Przasnyski, 2002; Fornell, 2006).

As mentioned previously, many of these methods became popular and widespread, but the question in many cases remains: Did it actually work? An analysis of many of these methodologies and their impacts is directly analogous to this praxis.

Total Quality Management, for example, focuses on driving to improved Customer Satisfaction through better market orientation, delivering better value to customers, and being responsive to changing marketplace needs while improving efficiency by reducing rework and reducing cost of conformance. The expected outcomes are increase in sales,

market share, and profits (Hietschold, 2014). Research regarding TQM and financial performance used surveys and interviews to measure performance by collecting opinions about financial performance utilizing a Lickert scale (George, 1998; Anderson J. R., 1995; Adam, 1994). Other analyses found no impact to firm performance through the use of TQM (Yunis, 2013; Wayhan, 2007). The most comprehensive and among the only research to use real financial data used T tests to show that organizations with award winning implementations of TQM performed slightly better in terms of cost based measures and OER and slightly better in terms of Return on Sales (Hendricks, 1997). Yet, a later analysis showed that organizations that performed well also performed well prior to implementation of TQM or the receipt of award, while under performing organizations still underperformed after the implementation of TQM (York, 2004). It was found that there is no evidence that the performance of successful firms improved due to the implementation of a quality management program (Zhang G. P., 2012).

Six Sigma is another methodology widely promoted and adopted but with very mixed results in empirical research. One study found that Return on Assets improved through improvement in operational efficiency and reduced costs with the use of Six Sigma projects, but also found that the benefits were significantly correlated with financial performance before its adoption (Swink, 2012). Other research indicates that Six Sigma may actually have negative impacts. 91 percent of large companies that announced Six Sigma programs trailed the S&P 500 since, though most of the strongest criticism is that it stifles true innovation (Morris, 2006; Bogle, 2008). Through the 2000's, rigorous empirical research regarding firm level impact began to emerge, but contradicted the positive anecdotal evidence, as multiple studies showed no significant main effects in

terms of Return on Assets, Return on Investment, total assets, asset turnover, or cash flow per share (Foster, 2007; Shafer, 2012). Reuters showed that leading Six Sigma companies did not outperform the stock market as a whole

ISO implementation is another good analog. In a similar longitudinal study to this one, operating performance (the ratio of operating profit to revenues)was measured over the course of 5 years, beginning one year prior to the implementation of ISO. A Wilcoxon Sign test was utilized to compare the median operating profit from year -1 (the year preceding ISO certification) with each of the following four years. This study showed a slight positive improvement in operating performance after ISO certification (Aba, 2016). Previous research into the question was mixed, with several studies that found implementation of ISO 9000 did not result in improved quality, productivity, or profitability (Corrigan, 1994; Lima, 2000). More nuanced research showed that companies that approached ISO certification with internal motivations to improve achieved positive firm level impacts while those that obtained certification to meet requirements for contract bids or due to pressure from customers did not (Woan-Yuh, 2008).

2.7.3 Agile Performance

Within the research on Agile methods, however, while there have been hundreds of case studies and articles, there has been little empirical work to show bottom line performance. For the bulk of the research, performance improvements are anecdotal and cited in individual case studies almost universally based off of surveys asking questions regarding reduction in time to market, increased velocity, and improved quality (Rico, Sayani, & Sone, The Business Value of Agile Software Methods, 2009). Moreover, there

have been very few peer reviewed articles that address agile performance at any level, though there have been some high quality industry reports such as the QSMA report discussed below (Quantatative Software Management Associates(QSMA), 2008). The vast majority of case studies and data come from consultants in the Agile business, or are accessible through Agile focused organizations like Scrum Alliance and Scrum inc.

Previous studies on performance effects of Agile methods have measured intermediate impacts rather than bottom line impacts (Rico, Sayani, & Sone, 2009). As such, it remains unclear whether the implementation of these frameworks would ultimately result in a competitive advantage.

For example, one of the most comprehensive studies on Agile performance metrics to date established that firms were 37 percent faster delivering software to market, 16 percent more productive, and able to maintain normal defect counts despite schedule decompression (Quantatative Software Management Associates(QSMA), 2008).

Similar studies have been conducted and have used polling data to show that experts considered agile methods to be an improvement over traditional methods in terms of cost, quality, project success, productivity, job satisfaction, cycle time, communication, and time to market (Ambler S. , 2008; Version One, 2016; Ghani, 2015; Rico D. H., 2009). It has been noted that "Little research has empirically examined the software development agility construct in terms of its dimensions, determinants, and effects on software development performance" (Lee, 2010). Productivity metrics are embedded to some degree within many methodologies but are difficult to compare across teams, much less organizations (Downey, 2013).

That said, a systematic literature review evaluated 274 articles relating to Agile development, but only 28 of those articles provided any data to establish a link to improved operational productivity, as well as some relation to client satisfaction, quality, and employee motivation (Cardozo, 2010).

Other studies have reaffirmed the link between use of agile methodologies, productivity, and project success, but not firm financial performance, which is evaluated in this paper (Tonelli, 2013; Quantatative Software Management Associates(QSMA), 2008).

The most extensive quantitative research to date has been conducted by Dr. David Rico, who evaluated hundreds of studies on Agile Methods and identified 79 that had data that could be extracted that was informative regarding Return on Investment in Agile methods. This research focused extensively on specific technical practices, was reliant on project data, and again, was not informative at the firm level. While the findings were compelling, "the final verdict on the cost and benefits of agile methods has not been reached" (Rico D. H., 2009; Rico D., 2008). An extension of this research established a link between some Agile practices and website quality (Rico D., 2007).

The research has not all been positive. A study of 8 Russian software companies using data from 35 projects found that schedule and cost performance decreased, though quality increased (Suetin, 2016). Conversely, an Australian case study showed significant productivity gains with the implementation of Scrum (Kautz, 2014).

Research on performance of specific methodologies other than scrum is very difficult to find. A recent systematic literature review evaluated 3,242 articles between 1990 and 2012 that were related to Kanban. Of those, only 37 had information regarding the

positive effects of the implementation of Kanban. Of the 37 articles, only 7 were in peer reviewed journals, and 8 in books. The remainder came from web articles, theses, and conference presentations and proceedings. The studies reporting positive results did not report enough quantitative data to evaluate empirical performance improvement, but did capture that the largest benefits were enhancement of visibility to facilitate decision making, assisting the coordination of cross functional teams, introducing quality improvement initiatives, reducing cycle time, increasing customer satisfaction, build high performing teams, enhancing quality, and driving organizational change (Al-Baik O. a., 2015).

Quantitative comparison between Agile frameworks was also difficult to find. One study compared the team level productivity of their Scrum teams, and again when they transitioned from Scrum to Kanban to their Kanban teams. After transitioning to Kanban, they found an increase in overall productivity of 21% and their cycle time was halved, with comparable quality results (Johnsen, 2012).

It is clear that there is a dearth of empirical data regarding Agile performance, especially at the organizational level.

2.7.4 Challenges in measuring firm level performance

Research regarding factors that impact firm level performance is popular and widespread, but even where positive results are shown, the effect sizes are often minimal (Murphy, 2016). When trying to measure performance at the organizational level, March (1997) identifies three primary reasons why studies evaluating firm level performance are often inadequate. The three factors he noted are instability in performance advantage, use of over-simplified models, and challenges of retrospective recall. Some models have

attempted to mitigate these issues using large sample sizes, but this may not be sufficient (Duarte, 2011). Another significant challenge is in identifying and measuring factors within the organization that can impact firm performance.

First, there is significant difficulty in identifying and measuring organizational factors. Data is obtained through using polls that ask employees their perception on improved quality, productivity, time to market, and profitability. While this is a widely used method, it is difficult to measure operational practices without direct observation and empirical data (Hansen, 1989).

As noted above, March (1997) cites retrospective recall as a major issue in evaluating causes of firm level performance. The vast majority of research using a measure of firm performance as the dependent variable utilizes retrospective accounts as the source of data. Polling data is generated by asking about empowerment, engagement, group cohesiveness, and changes to those factors over time. These studies are particularly vulnerable to retrospective bias (March, 1997). It was also shown that perceptions of firm quality were "more closely related to prior financial performance than to subsequent financial performance (McGuire, 1990). In this study only publicly available financial data that is subject to accounting regulations is utilized to calculate the dependent variable. The transition year is also factual data not based on polling, so this concern is addressed.

Instability in performance advantage is another factor that makes firm level performance research difficult. Performance instability exists because the business environment in which firms compete is dynamic and there is a significant level of competitive imitation that occurs.

- Any activities that may constitute competitive advantage are often copied and thus progressively eliminated.
- This 'institutional diffusion' reduces the variation in effective methods and ends up obscuring the effects.
- Not all of the institutional diffusion is captured in firm documentation, so researchers are not often aware of the potential dilution of competitive advantage due to imitation.

This effect has been the most widely used explanation for the relatively poor performance of operations research in the prediction of firm level performance (March, 1997).

In the case of implementing Agile methodologies and the purposes of this study, this effect is mitigated in large part because a complete overhaul of team structures, reporting, and operating mechanisms is usually required.

Using simple models for complex interactions is the third issue raised by March (1997). He is particularly critical of cross-sectional studies. Where all measurements are taken at the same time, the choice of what factors are causally dependent is difficult to show. Performance is also often strongly correlated with prior performance, as are many of the factors that might impact future performance.

The nature of this study mitigates many of the concerns identified above. Because these transformations are a radical departure from previous operating models, evaluation of each firm's before and after performance offers an opportunity to evaluate the full impact of making the change.

Because the financial data utilized is subject to financial reporting regulations and the year in which the transformation occurred is not based in opinion, this study is not subject to retrospective bias.

2.8 Statistical Methods

While this study utilizes statistical techniques that are well established and accepted, there are some techniques that are not traditionally used in systems engineering or engineering management. For that reason, a brief description of Repeated Measures ANOVA and Change Point Analysis follows.

2.8.1 Longitudinal Data Analysis

Traditionally, observational studies of this type are cross sectional in nature. Cross sectional studies compare different groups at the same point in time. For instance, if you wanted to evaluate cholesterol levels you could look at cholesterol levels, demographic data, and fitness level of all participants at the same time. Often correlation analysis is performed, but this does not provide definitive information about cause and effect relationships (Barkaui, 2014).

Longitudinal analysis, on the other hand, is a type of observational study that observes the same subjects over a period of time, allowing for the detection of changes at both the group and the individual level. The benefits of longitudinal analysis are well documented. First, it is widely accepted as providing a better basis for claims of causality than cross-sectional studies because the temporal order of cause and effect variables is known. It also allows visibility into change over time. Yet, cross-sectional studies predominate, in part because longitudinal data is generally much harder to come by (Barkaui, 2014).

2.8.2 Repeated Measures ANOVA

Repeated Measures ANOVA is among the most widely used statistical techniques in neuroscientific, psychological, medical, agricultural, and social scientific fields. Organizational research has been increasingly utilizing multilevel modeling techniques. A recent survey of the Journal of Applied Psychology, Personnel Psychology, and Organizational Behavior and Human Decision Processes indicated that of over 600 articles, over ten percent utilized either Repeated Measures ANOVA, Multivariate Repeated Measures ANOVA, or Repeated Measures regression (Misangyi, 2006). The same study shows that Repeated Measures Regression, despite its relatively wide adoption, is suitable for only a small number of situations and that for designs where between-subjects factors are limited to group membership, as in this study, the univariate RM ANOVA is the most appropriate, though if the data is unbalanced a Multilevel Modeling approach may be necessary (Misangyi, 2006).

A brief survey of the George Washington University dissertation database showed several studies where RM ANOVA was the primary research methodology. For example, an RM ANOVA was used to compare the outcomes from technology investment evaluation methods that included Decision Trees and Real Options (Wang, 2007). It was also the analysis used to assess the effectiveness of a computerized working memory intervention on math achievement, fluid reasoning, and learning constructs where the subject data was obtained through data regarding ADHD diagnosed children (Heishman, 2015).

In general terms, whether the approach is RM ANOVA, Linear Mixed Models, or General Linear Model – Repeated Measures, the distinguishing feature of this methodology is the use of longitudinal data with primary focus on within-subjects effects.

Within-subjects designs are best suited for measuring the change of outcome over time, and each subject becomes their own 'control'. In within-subjects designs, the within –subjects factor indicates that the same participants are measured on the same dependent variable on the same time points. Each within-subjects factor has categorical levels, and multiple within-subjects factors can be assessed. In such cases, one of the independent variables is considered a focal variable, and the remaining independent variables are moderator variables. In time series longitudinal analyses, time is the focal variable that moderates the effect of the other within-subjects factor.

Repeated Measures ANOVA is considered significantly more powerful statistically. Within-subjects designs are more statistically powerful (Seltman, 2015).

> "We can partition the variance due to individual differences from the rest of the "error" variance. Thus, the total variance in the within-subjects ANOVA is comprised of treatment variance, between-subjects variance (i.e., individual differences), and error variance. We still determine the effect of the treatment by examining the proportion of treatment variance to error variance. By partitioning out the between-subjects variance, we reduce the amount of error variance in the equation, thus reducing the "noise" we have to see through in order to see a significant treatment effect. Put another way, since we are not interested in differences between participants in a within-subjects design, we can throw out the between-subjects variance to get a clearer picture of what is going on in the data. (David, n.d.)."

We know that differences in means must be due to the treatment, the variations between the subjects (in this case firm size, firm age, etc.), and error. Essentially, by using multiple measurements for each subject (usually over time), the variability due to other factors such as subjects age, health, environmental factors is avoided, because each

subject acts as its own control. In other words, any factor that may affect the dependent variable will be exactly the same for the different conditions because they are the same subjects in the conditions (Hall, n.d.). As such, relatively minor differences within each subject can be detected despite much larger differences between the subjects (Lane, n.d.). For the purposes of this study, however, economic factors must be controlled for because not all firms were measured over the same 8 year time period. Substantial environmental factors could impact the performance of a number of organizations. That said, environmental factors such as firm size and age are likely to be significantly less impactful than in a cross sectional study.

It is important to note that the RM ANOVA and related tests are omnibus tests. The RM ANOVA will tell you whether the means are the same or not, but not what means are different. To get that information, post hoc testing is necessary. The two primary ways of doing so are using complex contrasts or capturing pairwise data (both with Bonferroni adjustment). Pairwise data is the more straightforward method. The difference in means is calculated between each possible pair of time points. By doing so, you can see which values show a statistically significant difference from each other. Contrasts, on the other hand, involve averaging the results of two or more treatments for comparison.

2.8.3 Change Point Analysis

Change Point Analysis is a relatively new technique that has proven to be a powerful statistical tool for identifying whether a change in the mean of time series data, and if so, when the shift occurred. It was developed to bolster analyses in Statistical Process Control (SPC). Typical use is to perform change point analysis on cumulated data periodically to detect changes too subtle to show up in control charts or to better

characterize the timing and nature of changes identified in control charts (Taylor W. , 2000). It has been used widely in the analysis of time-ordered data and identifies that a change to the mean has occurred and the time at which the change occurred (Gavil, 2009). Prior to the use of Change Point Analysis, the dominant method was to produce a CUSUM chart (cumulative sum) and to interpret the data visually, but CUSUM charts rely on visual inspection of the plot and can only detect large changes while not being reliable at identification of the actual time at which a change began (Gavil, 2009).

The benefits of Change Point Analysis are as follows (Taylor W. , 2000):

- It is a powerful way to detect relatively small sustained changes
- Reduces false detection by controlling the change-wise error rate.
- Robust to outliers.
- Can provide confidence levels and detect multiple changes
- Flexibility to multiple types of data, including attributes, individual values, counts, averages, and standard deviations
- Easy to interpret

In order to conduct Change Point Analysis, it is necessary to first construct a CUSUM chart (this would display the cumulative sum of differences between individual values and the mean). Traditionally, CUSUM charts would be used to evaluate change to the mean, but only relatively large changes can be identified. A sharp change in the direction of the CUSUM chart would indicate a possible change to the mean, but interpretation is subjective (Taylor W. 2000).

Change Point Analysis builds on the plotted CUSUM chart utilizing a bootstrapping approach. Essentially, each bootstrap generates a random iteration of the existing data

set. Each time this happens, there is another set of cumulative sums generated along with the difference between the highest and lowest CUSUM values. Then by finding the number of times the original CUSUM data exceeds the range for the bootstrap CUSUM data and expressing it as a percentage, you obtain the confidence level for whether a change to the mean occurred. Where other possible changes are present (as marked by changes in the CUSUM chart), data can be divided into subsets, thus multiple changes to the mean of the time series data can be detected simultaneously (Gavil, 2009).

There are two primary drawbacks to this analysis. It does not detect isolated abnormal points and the bootstrapping approach does not produce identical results each time it is performed due to the random selection of bootstrap samples. For example, The second issue is mitigated by using a large number of bootstraps (Taylor W. , 2000). The approach has been growing in usage and popularity and has been used in such wide ranging applications as pharmaceutical manufacturing and to investigate the wintertime ecophysiology and behavioral patterns of the raccoon dog (Mustonen, 2012).

Researchers using this tool have recently published articles in major journals, including The Impact of a Celebrity Promotional Campaign on the use of Colon Cancer Screening in Internal Medicine, Movement-Related Changes in Synchronization in Human Basal Ganglia in Brain, and 300 Hz Subthalamic Oscillations in Parkinson's Disease, also in Brain (Taylor W. , 2000; Cram, 2003; Cassidy, 2002; Foffani, 2003).

2.9 Summary of Literature Review

Agile frameworks are well documented and take significant effort to implement on a large scale. Likewise, their contribution to firm performance is critical in the justification of the effort involved in implementing Agile transformation. By using longitudinal

analysis, this study will be the first to address whether the implementation of these methods lead to improved firm level performance.

Chapter 3: Methodology

3.1 Experimental Design

The ongoing Agile movement within the Software Development sector offers a unique opportunity to perform causal analysis that mitigates the difficulty in building effective models, as discussed later. Because there are numerous case studies of agile transformations since the formalization of the Agile Manifesto in 2001, we can identify the point in time in which many organizations undertook a significant operational transformation. Because these transformations are radical departure from previous methods, evaluation of each firm's before and after performance offers an operational study that is unique in operations management literature and is more indicative in terms of showing causality.

By using longitudinal data, this study sidesteps many of the challenges in assessing causality within operations management research and provides a unique analysis to evaluate complex systems performance using a relatively simple model that has been used to great effect in neuroscience research (Misangyi, 2006).

Comparison between organizations using Agile and those using cross-sectional methods is difficult, because of the between-subjects effects, and thus very large sample sizes would be required. However, repeated measures designs offer far more statistical power with fewer subjects because these designs control for factors that cause variability between subjects. By using longitudinal data and getting firm performance before and after the implementation of Agile methodologies is started, the subjects become their own controls because the model will assess how each subject will respond to intervention (Frost, 2015).

This study uses a quasi-experimental approach. The experimental structure is as follows. Return on Assets (ROA), Operating Expense Ratio (OER), and Revenues are measured for each subject organization over the course of 8 years. In year 5, the traditional operating model is replaced with an Agile Framework. Each framework is treated as a between-subjects factor. Other between-subjects factors serve as a control. That said, the subjects of this study all implemented their Agile frameworks at different times, as identified in case studies, press releases, and conference presentations. Data was normalized, with year 5 as the transition year.

3.2 Measures

3.2.1 Dependent variables

Firm financial performance is measured in a variety of ways. These are organizational level metrics (measures assessed on the organization financial documentation) that are used to evaluate the overall health of an organization, its profitability, and whether it is worth investing in. Most studies focus on one or two measures, though analysts and investors tend to look at several firm level metrics when doing a full analysis of the organization. Top line measures would be things that appear near the beginning of a financial statement; things like revenues or gross sales and typically are measures of gross income. The bottom line is generally seen as net profit, and is often related to top line performance. Profitability metrics measure efficiency or return per company size; Operating Expense Ratio (OER), Return on Assets (ROA), and Return on Equity(ROE) fall into this category. The most commonly used measures are Revenues, ROA, and ROE, though other metrics are used less frequently (Rico D. H., 2009).

Agile frameworks attempt to improve operating efficiency, throughput, quality, customer satisfaction, reduced overhead, greater alignment with business priorities, and shorter time to market. Greater alignment between operational focus with business priorities along with reduced time to market and higher customer satisfaction should lead to increased top line performance, but may have a significant lag time associated with improvement. Because only some of the expected benefits of Agile implementation are expected to impact top line performance, efficiency and profitability metrics like OER are the most likely to be impacted with less lag time than top line performance metrics.

In this study, the most critical measure identified is the Operating Expense Ratio (OER) as it is a measure of profitability and efficiency. Operating Expense Ratio is the Operating Expenses/Revenues (Investing Answers, n.d.). Thus, the lower the OER the more efficiently the organization is generating revenue.

Revenues is the top line measurement of raw income a company generates from its costs and services. For the purposes of this study, to minimize the impact of significant differences in size of organizations, a ratio was used comparing the revenues in a given time frame to the revenues generated in the transition year. Thus, for the transition year, the revenue ratio will always be 1. Additionally, all revenues were adjusted for inflation and set to an equivalent in 2017 based on data from the Bureau of Labor Statistics (BLS) (Bureau of Labor Statistics, 2017). For foreign based organizations, inflationary data was captured using the Trading Economics website for country specific data (Trading Economics, 2017). This will control for inflationary effects over the course of the study.

Return on Assets is just that, the amount of profit generated expressed as a percentage of its total assets. This is widely considered as the best firm level metric for

investors and researchers, as it measures the overall profitability of the organization. Return on Equity is a similar metric that measures profit generated as a percentage of total shareholder's equity. ROE is also widely used in Operational research, but is much more volatile than ROA as it is particularly vulnerable to cost and debt structures, write downs, and share buybacks, which can artificially boost ROE (Investing Answers, n.d.). As such, this study measures ROA, not ROE. Higher ROA and ROE are associated with higher profitability and efficiency.

By utilizing Revenues, ROA, and OER a solid picture of overall firm performance becomes available. Revenue growth captures top line growth, while OER directly measures operational efficiency. ROA is a direct profitability metric that is a balance between Revenue and OER measurements.

3.2.2 Independent Variables

As a repeated-measures longitudinal study, it is necessary that all independent variables are clearly identified and understood as within-subjects factors or between-subjects factors (also known as treatments). In this case, as with most repeated measures designs, the only within-subjects variable is time. For each subject, 8 measurements were taken, at 1 year intervals. Years 1-4 were pre-transition data, and year 5 was the year in which the transition occurred, while 6-8 represent post transition data.

Because this study also seeks to identify the magnitude of the differential between the implementation of different methodologies, performance was also evaluated based on whether the organization implemented Scrum, Kanban, SAFe, Scrumban, LeSS, or DaD.

Because there are many factors that could impact the financial performance of target firms, it is necessary to identify and control for the factors most likely to impact the

validity of the study. To a great extent, external variables are controlled for by the nature of the study. Factors inherent to each subject in the study remain the same as it is the same subject being tested in each condition, so the effects of differences in each subject can be excluded (Field, 2011; Howitt, 2011). As such, running a one-way repeated measures analysis of variance is likely sufficient. The primary limitation in repeated measures designs is order effects, which are not directly applicable to this study (Owen, 2011).

That said, because the study spans several years, and the overall economic environment can have significant impact on firm performance throughout the industry, the greatest weakness in this analysis is the strong dependence on overall market financial conditions to firm performance. This is mitigated in part by sampling firms across a very large time frame, with a span from 1996 to 2015. To that end, evaluating periods of recession and identifying those as an Economic Environment factor if the transition dates were at or near the actual recession period is a critical control. Measures from years in which a recession was present and for one year after are identified as impacted by the economic environment and classified as a Bear market. Otherwise, the economic environment was classified as Bull.

Firm size can also greatly impact performance characteristics of organizations and could potentially impact the analysis. As such, the firms were categorized by their Market Capitalization size as Small (Market Capitalization less than $1 Billion), Medium (Market capitalization between $1 Billion and $4 Billion), Large (Market Capitalization of $4 Billion to $200 Billion), and Mega (Market capitalization greater than $200 Billion). It should be noted that there are no official definitions of Market Capitalization

Size, and that these values change over time. Size restrictions for inclusion in funds include some overlap: to be on the S&P 500 Large Cap Index, a company must have at least a $4 Billion Market Cap, while to be on their MidCap400 and SmallCap600 a firm would need to have Market capitalization between $1 Billion and $4.4 Billion and between $300 Million and $1.4 Billion respectively (Merrit).

Other factors are incorporated into the more sophisticated model. Firm age was identified to differentiate performance between startups and long-established organizations, and categorized as up to 5 years, between 5 and 10 years, between 10 and 20 years, and over 20 years. Firm geography focused on the bulk of firm operations and headquarters and divided into US, UK, Eurozone, and Korea. Firms were also classified according to industry.

Table 3-1 Summary of dependent and independent variables.

Variable Name	Type	Description
OER	Dependent	Operating Expenses expressed as a percentage of Revenues
ROA	Dependent	Net profit expressed as a percentage of total Assets
Revenue Ratio	Dependent	Revenue ratio as compared to year of transition
Time	Independent	measures taken annual for the duration of the study
Economic Environment	Independent	Whether a recession was in place or occurred within a year of the transition year
Firm Size	Independent	Firms categorized based on Capitalization size, from Small to Mega
Firm Age	Independent	Range from startup to over a century old
Industry	Independent	differentiates firms by specific industry
Geography	Independent	US, UK, Eurozone, Korea

3.3 Sample and Data Collection

The following criteria were required for firms to be considered appropriate for this study:

- Clear identification of the year in which a transition from traditional to agile methods occurred
- Traditional methods had to be dominant before transformation: Agile methods had to be dominant after
- Transformation had to directly impact the majority of the organization
- Publicly available financial information needed to be available that met minimum accounting practice guidelines

Organizations were selected for this study by first identifying organizations that had transitioned from using traditional methodologies to using Agile methods. This was done through a methodical search of press releases, case studies, and journal articles. A detailed evaluation of the organization was then performed through research on their website, financial reports, and other media to assess whether the bulk of their operations were utilizing agile methodologies. In some cases, the year in which the transformation occurred could not be immediately verified, in which case we contacted the author of the case study for further information. The Agile framework utilized was recorded and the availability of financial data confirmed. Note that this was a very time-consuming process. Each case study had to be evaluated thoroughly, with significant additional firm level research. Annual financial data was obtained annual for a total range of 8 years. The first four years represent data prior to the transformation, year five was the transition year, and years six through eight represent post transition data. The year in which large scale transformation began was considered to be after the transition for evaluation purposes.

The reasons that firms were rejected from this study are as follows:

- Transformation impacted only the IT portion of a non tech organization
- Transformation occurred in a single division of a multiple division organization
- Date of transformation was not able to be verified
- The organization had insufficient available financial data

3.4 Study Design

3.4.1 Difference Before and After

For an initial analysis of the impact of transition, for each measure a paired T test was performed to compare the mean of years 1-4 to the mean of years 5-8. Paired T tests are used to evaluate data before and after where participants are the same individuals (Mowery, 2011). Where the assumptions for the paired T test are not met, the Sign test was used.

3.4.2 Repeated Measures ANOVA (RM ANOVA)

To provide greater clarity, this was followed by a longitudinal analysis using the General Linear Model function in SPSS using Bonferroni adjustment and using complex contrasts to evaluate the main effects (if any) (Grace-Martin, n.d.).

For this study, a mixed Repeated Measures ANOVA was utilized. This was accomplished using the General Linear Model (GLM) function in SPSS and choosing the Repeated Measures design. There were 8 levels identified (4 years before and 4 years after) and the test was repeated for each measure (ROA, OER, and Revenues). A Mixed Model approach was used, identifying the previously mentioned covariates to identify where significant effects are present (UC Denver; Taylor A. , 2011).

3.4.3 Change Point Analysis

Change Point Analysis is used to identify if there has been a shift in the mean of time series data. Use of Change Point Analysis will identify whether or not there has been a shift in the mean and also identify at which point in time the shift occurred. For the purposes of this study, it is expected that there will be a change identified during year 5 for ROA, OER, and Revenues.

For this analysis, the Change Point Analysis tool from Taylor Enterprises was used. This tool allows for the use of multiple observations per time period and provides easy to understand charts and tables identifying whether a change occurred and at what point the change occurred.

3.4.4 Chow Test

To calculate the Chow test, linear regression was performed on the entire dataset for OER, ROA, and Revenues. Regression was then repeated for before and after transition data for all dependent variables and the F statistic was calculated using equation 3-1, where RSS_P represents the combined regression line, RSS_1 is the residual sum of squares before the break, and RSS_2 is the residual sum of squares after the break. k is the number of estimated parameters and N_1 and N_2 are the number of observations in the two groups.

$$F = \frac{(RSS_p - (RSS_1 + RSS_2))/k}{(RSS_1 + RSS_2)/(N_1 + N_2 + 2k)} \qquad \text{(Equation 3-1)}$$

Chapter 4: Results

4.1 Introduction

The results of the study are presented in the order in which they were performed. First, descriptive statistics on the dataset are provided, followed by preliminary screening procedures required for T testing, RM ANOVA, and the Chow test. Section 4.4 shows the results of the T testing, RM ANOVA, Sign test, Change Point Analysis, and the Chow test.

4.2 Descriptive Statistics

A brief summary of the descriptive statistics is located in Table 4-1.

Table 4-1 Frequencies and Percentages for the Company Variables

Summary of Descriptive Statistics					
Variables	**n**	**%**	**Variables**	**n**	**%**
Agile Methodology			**Industry**		
Scrum	16	51.6%	Software	9	29.0%
SAFe	9	29.0%	Business Services	8	25.8%
DAD	1	3.2%	Retail	2	6.5%
LeSS	1	3.2%	Telecom	3	9.7%
Scrumban	1	3.2%	Consumer Electronics	3	9.7%
Kanban	3	9.7%	Banking and Finance	2	6.5%
			Industrial, Construction, Heavy		
Size			Equipment	3	9.7%
Small	10	32.3%	**Geography**		
Mid	12	38.7%	US	21	67.7%
Large	6	19.4%	UK	3	9.7%
mega	3	9.7%	EU	4	12.9%
Age of Firm			Multinational	2	6.5%
less than 5 years	0	0.0%	Korea	1	3.2%
5-10 years	4	12.9%			
10-20 years	6	19.4%			
Over 20 years	21	67.7%			
Economic Environment					
Bull	25	80.6%			
Bear	6	19.4%			

4.3 Preliminary Screening Procedures

4.3.1 Assessing Normality and Outliers – General Approach

Outliers were evaluated using box plots for all analyses. Data points greater than 1.5 box lengths from the box edge are classed as outliers, while those more than 3 box lengths away are classed as extreme outliers and are labelled with an *. The recommendations for dealing with outliers from Laerd Statistics are as follows (Laerd Statistics, 2015): The first concern is to verify that it is not a data entry or measurement error. Assuming the value is correct, the following options are valid and acceptable:

1. If you feel you cannot remove an outlier, use a nonparametric test (Wilcoxon signed-rank test, sign test, or Friedman test).
2. Modify the outlier by replacing its value with one less extreme. This is not a widely used option because there are significant risks involved.
3. Transform the dependent variable. This is recommended only if normality is also an issue.
4. Keep the outlier in the analysis because you don't believe its inclusion will materially affect the result.

In regards to option 4 above, Laerd has this to say in regards to both the paired Ttest and RM GLM (Laerd Statistics, n.d.):

> "… keeping the outlier in the analysis requires a lot more confidence on your part, but can be a perfectly acceptable strategy in dealing with outliers. Ideally, you are looking to find a method that evaluates whether the outlier has an appreciable effect on your analysis. One method you can use is to run the test with and without the outlier(s) included in the analysis. You can then compare the results and decide whether the two results differ sufficiently for different conclusions to be drawn from the data. If the conclusions are essentially the same (e.g., both result in a statistically significant result, confidence intervals are not appreciably different, etc.), you might keep the outlier in the data."

For the RM ANOVA and the paired T tests, the Shapiro-Wilk test was used to assess normality. This was done using the Analyze/Explore function in the IBM SPSS software package (Laerd Statistics, 2015). Again, there are four ways to handle the deviation.

The data can be transformed or the nonparametric test can be run. Transformation of data and running the analysis on both transformed and the original data, and if the conclusions are the same, utilize the analysis from the original data. The last option is to "run the test regardless because the one-way repeated measures GLM and paired T tests are fairly "robust" to deviations from normality. Indeed, if sample sizes are not small, even somewhat skewed distributions – as long as the levels of the within-subjects factor are similarly skewed – are not always problematic. In conclusion, non-normality does not affect Type I error rate substantially and both the one-way repeated measures ANOVA and paired T test can be considered robust to non-normality. (Laerd Statistics, 2015)"

Where either outliers or normality is a problem, Laerd Statistics (2015) holds that the worst option is generally to remove potentially valid data points and generally recommends transformation. As such, where assumptions are not met, non-parametric tests are run as well as the planned, unaltered analysis. Where there is a discrepancy, it is called out. This allows for validation of the more powerful model if in agreement with the non-parametric test (Laerd Statistics, 2015).

4.3.1.1 Paired T Test

For dependent or paired sample T testing, there are four assumptions.

1. One dependent variable measured continuously.
2. One independent variable that has two categorical groups.
3. No significant outliers.

4. Distribution of the differences in the dependent variable between groups is approximately normally distributed.

The first two are met by the nature of the data, as there are two categories, before and after, that are being evaluated, and the test is repeated for OER, ROA, and Revenues.

For OER, there was not a normal distribution in the differences as the Shapiro-Wilk p values were less than 0.0005 for all scenarios. There were 2 outliers. The non-parametric Exact Sign test was utilized. A natural log transformation was used but did not result in a normal distribution and the same outliers remained. As such, only non-transformed data was used. The outliers for the before and after data were replaced with the next highest values which resulted in elimination of outliers. The non-parametric exact sign test was used. Additionally, because "non-normality does not affect Type I error rate substantially and the paired-samples t-test is often considered robust in this regard" and there is a moderate sample size, the parametric dependent T test was run as well. (Laerd Statistics, n.d.).

ROA did not show a normal distribution with a Shapiro-Wilk p of less than 0.0005. There were three outliers identified. As such, the Exact Sign test was utilized. Transformation using a natural log function was utilized by adding 1 to eliminate negative values. Transformation did not result in a normal distribution. As such, only non-transformed data was used. The outliers for the before and after data were replaced with the next highest values which resulted in elimination of outliers. The non-parametric exact sign test was used. Additionally, because "non-normality does not affect Type I error rate substantially and the paired-samples t-test is often considered

robust in this regard" and there is a moderate sample size, the parametric dependent T test was run as well. (Laerd Statistics, n.d.).

Analysis of the distribution of the differences for Revenue Ratio was more straightforward. Using the Shapiro-Wilk method showed the data was normally distributed. There were 3 outliers. Because transformation is recommended only when the assumption of normality is violated, the outliers were altered to the next most extreme data and the paired T test was run (Laerd Statistics, n.d.). For completeness, the exact sign test was completed for this test as well.

4.3.1.2 GLM Repeated Measures

In a Repeated-Measures GLM, there are five assumptions that must be met.

1. There is one continuous dependent variable.
2. The within-subjects factor is categorical and has at least three levels.
3. There are no significant outliers in any level of the within-subjects factor.
4. The dependent variable is approximately normally distributed at each level of the within-subjects factor.
5. Variances of the differences between levels of within-subjects factor are equal. This is known as sphericity.

Assumptions 1 and 2 are met by the nature of the data, as each dependent variable is a continuous variable and had a within-subjects factor (independent variable) that represented the before and after transformation measurements (Laerd Statistics, 2015; Singh, 2013; Tamura, 1992).

OER was normally distributed as assessed by the Shapiro-Wilk's test($p > 0.05$) for all levels of the data. OER also had no outliers, meeting the requirement for the Repeated Measures GLM test.

ROA was normally distributed at each time point except for the second and third years of the study as assessed by the Shapiro-Wilk's test with p values of 0.003 and <.0005 respectively. ROA also showed several outliers prior to implementation. In order to utilize a natural log transformation, it was necessary to add a constant of 1 to each value to eliminate negative values. Transformation resulted in more outliers and the treatments in years 2 and 3 remained non-normal. As such, non-transformed data was used. One extreme outlier was identified and that data point was removed from the analysis. The non-parametric Friedman test was utilized, and the RM GLM was run as well. This is acceptable because "non-normality does not affect Type I error rate substantially and the repeated measures GLM can be considered robust to non-normality" (Laerd Statistics, 2015).

Revenue Ratios were normally distributed for the first 5 of the 8 time periods and showed several outliers after implementation but none before. Applying a natural log function did not improve normality measurements and actually increased the number of outliers prior to implementation. As such, the non-transformed data was used. There were two organizations that showed extreme outliers that were removed from the study. The non-parametric Friedman test was utilized, and the RM GLM was run as well. This is acceptable because "non-normality does not affect Type I error rate substantially and the repeated measures GLM can be considered robust to non-normality" (Laerd Statistics, 2015).

Mauchly's test of sphericity evaluates whether the variances of the differences between the levels of the within-subjects factor (time) are equal (Laerd Statistics, 2015). This is expected, as in practice this assumption is difficult to meet and some studies recommend using the Greenhouse-Geisser correction in all cases (Maxwell, 2004). For all measures, Mauchly's test of sphericity was violated. As such, the Greenhouse-Geisser correction was used (Laerd Statistics, 2015).

4.3.1.3 Chow Test

In order to run a Chow test, a linear regression must be used. The assumptions for linear regression are as follows (Casson, 2014; Laerd Statistics, 2015):

1. The study must incorporate a continuous independent variable and a continuous dependent variable.
2. There must be a linear relationship between the dependent and independent variables.
3. There should be independence of observations.
4. Data must show homoscedasticity.
5. Residuals of the regression line must be approximately normally distributed.

Assumption 1 is met because the independent variable is time, a continuous variable. All dependent variables are continuous as well. All observations for OER, ROA, and Revenues are independent, which satisfies the third assumption.

The preferred method for evaluating the remaining assumptions is evaluation of graphical data, as outlined by Chambers (1983, p. 1) and codified in the statistical guidelines for the APA (Wilkinson, 1999). In fact, use of formal tests is strongly discouraged by many (Albers, 2000). For the purposes of this study, guidance on interpretation of plots was taken from Casson (2014).

To that end, a scatterplot of ROA vs. time was plotted. Visual inspection of the scatterplot indicated a linear relationship between the variables. This was repeated for Revenues and OER, and a linear relationship was confirmed for both.

OER and ROA both exhibit homoscedasticity (assumption 4) as assessed by visual inspection of a plot of standard residuals versus standardized predicted values. Revenues exhibited heteroscedasticity and cannot be transformed to alleviate the issue because year 5 is always normalized to equal 1. Any transformation would result in heteroscedasticity. The analysis was still performed, as "violations of the homoscedasticity assumption are not necessarily problematic. Provided that the very mild assumption of finite variance holds, estimates will still be unbiased and consistent (Ernst, 2017)".

Residuals for OER, ROA, and Revenues were all normally distributed as assessed by visual inspection of a normal probability plot and histogram.

4.4 Primary Statistical Analyses

It was hypothesized that operational expense ratios (OER) would be lower (first hypothesis), revenues would be higher (third hypothesis), and ROA would be higher (fifth hypothesis) after organizations implemented an agile methodology. It was also hypothesized that improvement in operational expenses (second hypothesis), revenues (fourth hypothesis), and ROA (sixth hypothesis) would differ as a function of type of agile methodology implemented.

Table 5-2 and Table 5-3 below summarize the results of the exact sign tests and paired T tests. It should be noted that the mean and median of the difference for the paired T test and the Sign test are calculated by subtracting the value after transition from the value before transition i.e.

$$Before - After \qquad \text{(Equation 4-1)}$$

As such, it should be noted that an increase in ROA or Revenues would result in a negative value. Likewise, a decrease in OER will result in a positive mean difference.

Table 4-2. Summary Results of the Paired T tests.

	Paired Differences					t	p
	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference			
				Lower	Upper		
ROA before - after	-0.07234	0.097089	0.017438	-0.107953	-0.036728	-4.149	<.0005
OER before - after	0.12371	0.228539	0.041047	0.0398868	0.2075444	3.014	<.0005
Revenue before -after	-0.38459	0.682304	0.122545	-0.634861	-0.134319	-3.138	0.004

Table 4-3 Exact Sign Test summary data.

Measure	Median Before	Median After	Median difference	# increase	# decrease	p	Reject Null
ROA	0.028	0.071	-0.04	5	26	<0.0005	Y
OER	0.86	0.8	0.066	26	5	<0.0005	Y
Revenue Ratio	0.72	1.13	-0.035	6	25	0.001	Y

The Exact Sign tests show a statistically significant difference in the median value for all Scenarios, with ROA and Revenues increasing after the advent of Agile methods and OER decreasing, as predicted. This is in agreement with the T tests, which show a statistically significant difference (improvement) in before and after performance for ROA, OER, and Revenues for all Scenarios.

This supports hypotheses 1, 3, and 5 that there was improvement in all three measures after the implementation of Agile methods.

The median values associated with the Friedman's test are summarized in Table 5-4. The Friedman's test recommended rejection of the Null Hypothesis that all median values are the same for ROA, OER, and Revenues.

Table 4-4 Friedman's test shows the median values for each measure at each time point.

Measure	*4 Years Before*	*3 Years Before*	*2 Years Before*	*1 Year Before*	*Transition Year*	*1 Year After*	*2 Years After*	*3 Years After*
ROA	0.022	0.031	0.021	0.043	0.078	0.052	0.065	0.056
OER	0.816	0.826	0.851	0.845	0.8	0.769	0.734	0.733
Revenue Ratio	0.671	0.8	0.734	0.881	1	1.12	1.18	1.24

The Repeated Measures ANOVA summary presented in Table 5-5 shows a statistically significant difference in the means, so the null hypothesis (that the means are the same) can be rejected. The model also shows no significant contribution to the change from any of the control variables.

The interaction between time and methodology for OER, ROA, or revenues, given by time*Method in Table 5-5 did not show a significant difference, indicating that the null hypothesis cannot be rejected. Thus we cannot say whether performance differs as a function of type of agile methodology. Thus, the second, fourth, and sixth hypotheses were not supported. Yet, some qualitative analysis can be done that may provide insight. Table 5-10 shows the results of a Sign test by methodology. Figure 5-5 shows the main effects plots for each method (Scrum, SAFe, Kanban) for ROA, OER, and Revenues.

Table 4-5. Repeated Measures ANOVA results from 4 years before Agile Transformation to 4 Years After for OER, ROA, and Revenue

Measure	Factor	Effect	df	F ratio	p	*η2*
OER	time	within-subjects	2.66	9.45	<.0005	0.291
	time*Method	within-subjects	6.35	0.89	0.49	0.08
	time*economic environment	within-subjects	1.22	0.66	0.52	0.248
	time*size	within-subjects	2.44	1.31	0.43	0.57
	time*age of company	within-subjects	2.44	1.55	0.38	0.61
	time* geography	within-subjects	1.22	2.25	0.26	0.53
	time*industry	within-subjects	2.44	1.43	0.4	0.59
ROA	time	within-subjects	1.83	7.62	0.002	0.248
	time*Method	within-subjects	3.98	2.35	0.07	0.18
	time*economic environment	within-subjects	1.282	0.43	0.599	0.125
	time*size	within-subjects	3.85	0.356	0.824	0.263
	time*age of company	within-subjects	5.127	0.274	0.908	0.268
	time* geography	within-subjects	1.301	0.29	0.688	0.127
	time*industry	within-subjects	1.28	0.312	0.661	0.094
Revenue Ratio	time	within-subjects	1.17	16.75	<.0005	0.401
	time*Method	within-subjects	2.37	1.14	0.34	0.09
	time*economic environment	within-subjects	1.125	0.284	0.66	0.124
	time*size	within-subjects	2.243	0.873	0.536	0.466
	time*age of company	within-subjects	2.243	6.001	0.126	0.857
	time* geography	within-subjects	1.12	0.392	0.612	0.164
	time*industry	within-subjects	3.364	0.638	0.664	0.489

note: η2 is an indicator of effect size. η2 > 0.14 is considered a large effect, and η2<0.06 is a small effect, df represents degrees of freedom and is comprised of two values. The first is degrees of freedom, followed by an error term. All values are using the Greenhouse-Geisser correction.

To further analyze the before-after performance, complex contrasts and pairwise comparisons were utilized with Bonferroni adjustment (Laerd Statistics, 2015). Table 5-6 summarizes this data. For each dependent variable, the complex contrasts compare the average of the means before the transition to each data point in years 5 through 8. For ROA, OER, and Revenues, this difference is statistically significant for each year, as noted by the p value of less than 0.05. The η2 value is significant for each measurement as well. It should be noted that ROA and OER show a relatively stable mean difference for each time, but the Revenue Ratio difference continues to grow.

Table 4-6. Summary of the Complex Contrast data.

		Avg. Before vs. year 5	Avg. Before vs. year 6	Avg. Before vs. year 7	Avg. Before vs. yr. 8
ROA	Mean Difference	0.073	0.072	0.072	0.063
	p	0.001	0.001	0.01	0.004
	$\eta 2$	0.4	0.37	0.26	0.31
OER	Mean Difference	-0.085	-0.098	-0.093	-0.089
	p	0.001	<0.0005	<0.0005	0.001
	$\eta 2$	0.36	0.44	0.43	0.37
Revenue Ratio	Mean Difference	0.25	0.41	0.59	0.79
	p	<0.0005	<0.0005	<0.0005	<0.0005
	$\eta 2$	0.41	0.42	0.42	0.41

note: η2 is an indicator of effect size. η2 > 0.14 is considered a large effect, and η2<0.06 is a small effect, df represents degrees of freedom and is comprised of two values. The first is degrees of freedom, followed by an error term. All values are using the Greenhouse-Geisser correction.

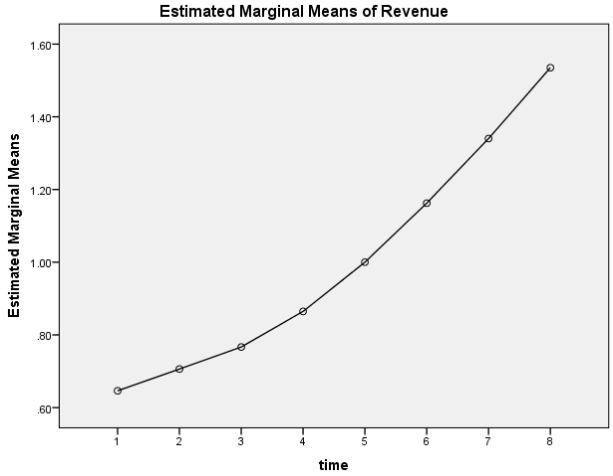
Post hoc testing data is presented in Table 5-7 and shows pairwise comparisons between each time point and each other time point. Where $p<0.05$ there is a statistically significant difference, with direction defined by the sign of the mean difference. Data can be understood by comparing time I with each of the time J rows. For example, time 1, the first year of the study can be compared to time 2 to show a non-statistically significant difference in the mean for ROA, OER, and Revenues with p values of 0.634, 0.710, and 0.155 respectively.

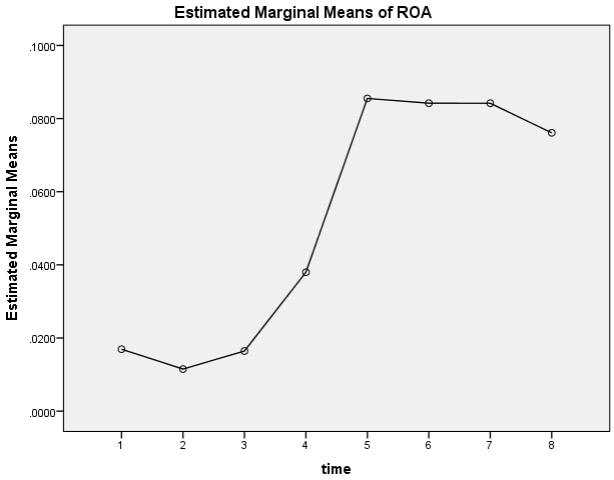
For OER and ROA, when you compare any of the values for time 1-4 with any of the values for time 5-8 there is a statistically significant difference in the means. Conversely, when you compare time 1 with time 2, 3, or 4, you do not get a statistically significant result.

Revenues also show clear improvement post transition over pre transition, but it should be noted that the only time periods that do not show a statistically significant difference are times 1 and 2. This indicates that the mean is changing significantly at almost every measurement. This is confirmed when looking at the main effects plots in Figure 5-1 which shows mean ROA, OER, and Revenue Ratio over time. The first 4 data points are prior to the roll out of Agile methods, the 5^{th} data point is the transition year, and 6-8 are the following years.

Table 4-7 Post Hoc pairwise comparisons of pre and post transition means.

		OER		ROA		Revenue Ratio	
Time I	Time J	mean difference I-J	p	mean difference I-J	p	mean difference I-J	p
1	2	0.013	0.634	0.005	0.710	-0.060	0.155
	3	0.007	0.652	0.000	0.973	-0.12	0.018
	4	0.024	0.229	-0.021	0.226	-0.218	0.002
	5	0.096	0.002	-0.069	0.004	-0.354	0.001
	6	0.109	<0.0005	-0.067	0.005	-0.516	<0.0005
	7	0.104	<0.0005	-0.067	0.023	-0.694	<0.0005
	8	0.1	0.000	-0.059	0.012	-0.889	<0.0005
2	1	-0.013	0.634	-0.005	0.710	0.060	0.155
	3	-0.006	0.758	-0.005	0.665	-0.061	0.024
	4	0.011	0.624	-0.026	0.104	-0.159	0.003
	5	0.083	0.007	-0.074	0.001	-0.294	<0.0005
	6	0.096	0.004	-0.073	0.002	-0.456	<0.0005
	7	0.091	0.006	-0.073	0.010	-0.634	<0.0005
	8	0.087	0.014	-0.065	0.006	-0.829	<0.0005
3	1	-0.007	0.652	0.000	0.973	0.12	0.018
	2	0.006	0.758	0.005	0.665	0.061	0.024
	4	0.016	0.192	-0.022	0.066	-0.098	0.001
	5	0.089	0.002	-0.069	0.001	-0.233	<0.0005
	6	0.101	<0.0005	-0.068	0.002	-0.396	<0.0005
	7	0.097	<0.0005	-0.068	0.010	-0.573	<0.0005
	8	0.093	0.001	-0.06	0.003	-0.768	<0.0005
4	1	-0.024	0.229	0.021	0.226	0.218	0.002
	2	-0.011	0.624	0.026	0.104	0.159	0.003
	3	-0.016	0.192	0.022	0.066	0.098	0.001
	5	0.072	0.005	-0.047	<0.0005	-0.135	0.001
	6	0.085	0.002	-0.046	<0.0005	-0.298	<0.0005
	7	0.08	0.003	-0.046	0.009	-0.476	<0.0005
	8	0.076	0.009	-0.038	0.001	-0.671	<0.0005





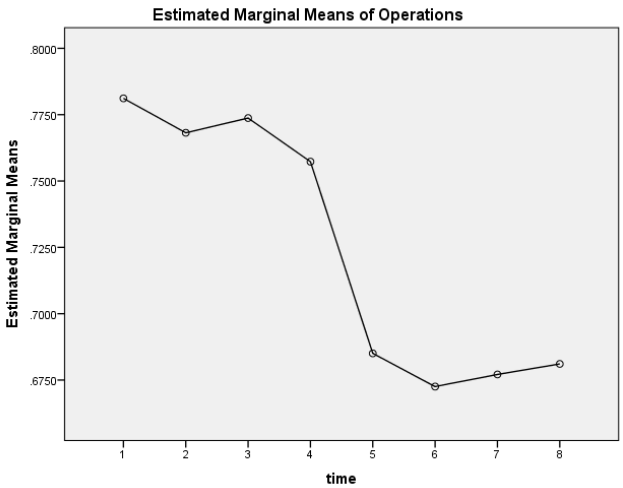


Figure 4-1. Main effects plots for Revenue, ROA, and OER over time.

Visual inspection of the Main Effects plots presented in Figure 5-1 show an apparent time effect in Revenue increases. OER shows relatively stable mean performance before transition and again after transition, with a discontinuity indicating improvement as a stepwise function. ROA shows a relatively flat performance rate after the implementation of Agile methods, but a possibly increasing rate prior to the implementation of Agile methods. The main effects plots qualitatively show behavior, but in order to fully assess the before and after effects measured in the GLM Repeated Measures test, Change Point Analysis was used, to be followed by a Chow test (Taylor W. , 2000).

In Table 5-8, Confidence Level shows the confidence level that a change to the mean level occurred. Confidence Interval identifies the time point or time frame during which there is a 95% level of confidence that the change to the mean occurred. In the table below, for both ROA and OER, the change to the mean was identified as having occurred during the Transition Year, as expected. Figures 5-2, 5-3, and 5-4 show graphical representations of the Change Point Analysis. Where the light blue portion shows a break, it indicates a change to the mean. For OER and ROA, this corresponds to the transition point. Revenues do not show a clear structural change to the mean.

Table 4-8 Summary of Change Point Analysis Data for OER, ROA, and Revenue Ratio.

Measure	# of Changes	Change Year (0 indicates trans. Year)	Confidence Level	Confidence Interval (95%)
OER	1	0	95%	(0,0)
ROA	1	0	95%	(0,0)
Revenues	n/a	n/a	n/a	n/a

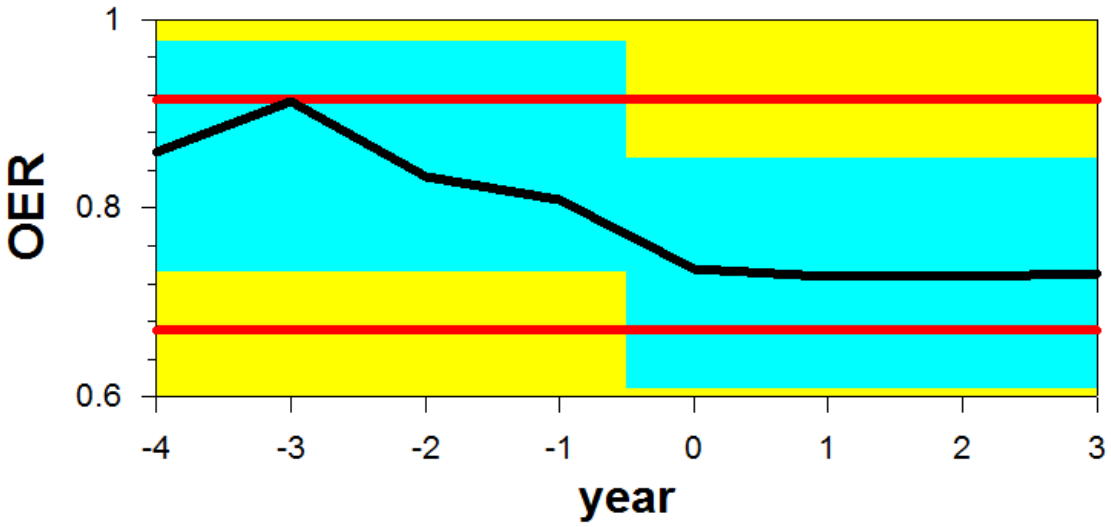


Figure 4-2 Graphical representation of the change using Change Point Analysis for OER. The blue highlights show the discontinuity at the point of change, which corresponds to the transition year.

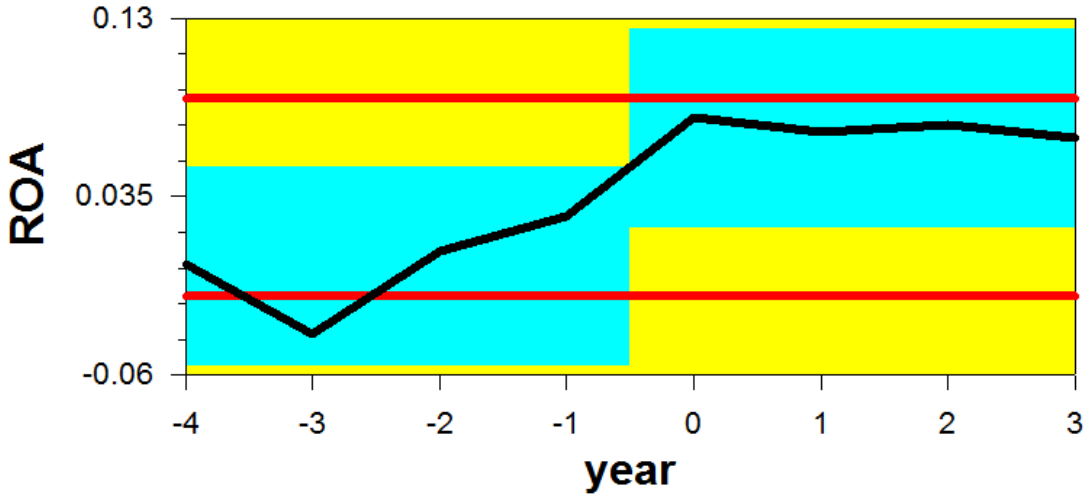


Figure 4-3 Graphical representation of the change using Change Point Analysis for ROA. The blue highlights show the discontinuity at the point of change, which corresponds to the transition year.

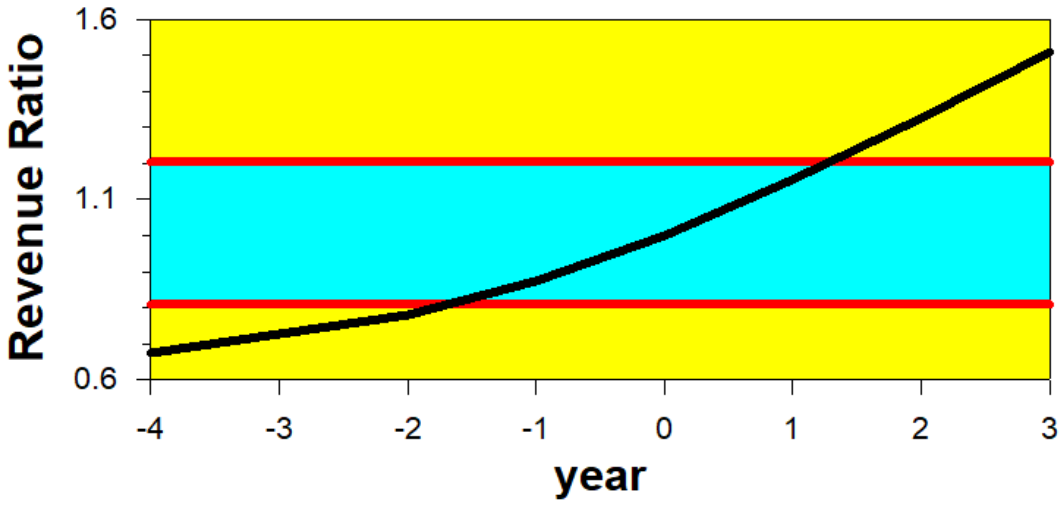


Figure 4-4 Graphical representation of the mean change using Change Point Analysis.

Change Point Analysis showed a clear statistically significant change in the mean during the Transition Year for both OER and ROA. The change is a discontinuity showing relatively consistent performance before the Transition Year and relatively consistent performance at an improved level after the Transition was begun. Revenues showed no change in the means for the data and indicates that improved Revenues after the implementation of Agile Methods are independent of the use of Agile methods.

While Change Point Analysis identifies break points in the data as well as identifying if there are structural changes to the data mean in time series data, to further test for structural change in the data at the time of transition, the Chow test was used. While the Chow test cannot detect break points, Change Point Analysis has confirmed that year 5, the transition year is the change point for both OER and ROA. The Chow test was performed assuming a change point of year 5 for OER, ROA, and Revenues. In the Chow test, the null hypothesis states that the relationship between the independent and dependent variables are the same between the before and after data. To put it another way, the coefficients of the regression model are the same across both groups.

To calculate the Chow test, linear regression was performed on the entire dataset for OER, ROA, and Revenues. Regression was then repeated for before and after transition data for all dependent variables and the F statistic was calculated.

F distribution tables for p=0.05 were used to evaluate whether structural change had occurred **(Dinov, 2012)**. An F statistic greater than the indicated level on the table means we reject the null hypothesis. Data for ROA, OER, and Revenues are summarized in table 5-9.

Table 4-9 Summary Chow Test data

Measure	df	F statistic	F critical value at p=0.05	Reject Null Hypothesis
ROA	2, 235	42.25	2.99	Y
OER	2, 237	3.77	2.99	Y
Revenues	2, 248	1.05	2.99	N

While the data in the paired T tests and exact sign tests points to acceptance of hypotheses 1, 3, and 5, a closer look at post hoc testing and main effect plots indicates that while Revenues were higher after the implementation of Agile methods, this was not likely due in significant part to the transition itself. This is confirmed through Change Point Analysis, which shows a clear change in both OER and ROA at the transition year, but no change point in the Revenue Ratio data.

Based on Complex contrasts there was a statistically significant increase in Revenue Ratio, OER, and ROA from the average of pre-transition to each of the years measured after. The effect sizes, $\eta_p 2$ are equivalent to R^2 values and represent the amount of variation is due to the temporal variation. $\eta_p 2$ values above 0.14 are considered significant (Laerd Statistics, 2015).

The lowest $\eta_p 2$ value above is 0.26. Thus, we can with confidence state that OER has improved (decreased) with the implementation of Agile methodologies with 40% of the variation explained by the transition. Likewise, ROA has increased (improved) with the implementation of Agile methods with 33% of the variation explained by the transition. Thus, we can conclude that hypotheses 1 and 5 are true: OER and ROA did improve after the implementation of Agile methods. We cannot confirm hypothesis 3, however, because Revenue increases are not causally related to the implementation of Agile methods.

We also have to reject hypotheses 2, 4, and 6, because the interaction effects between time and method were not significantly significant. A sign test was run for ROA, OER, and Revenues for organizations implementing Scrum, SAFe, and Other.

Table 4-10 Sign test by Agile framework

Measure	Method	Median Before	Median After	Median difference	# increase	# decrease	p	Reject Null
ROA	Scrum	0.044684	0.080779	-0.01484	13	4	0.049	Y
	SAFe	-0.02815	0.011974	-0.03946	8	1	0.046	Y
	Other	0.027774	0.083015	-0.16964	4	1	0.375	N
OER	Scrum	0.753807	0.659857	0.070717	3	14	0.013	Y
	SAFe	0.9012	0.861301	0.029761	2	7	0.18	N
	Other	0.861225	0.795882	0.104749	0	5	0.062	N
Revenue Ratio	Scrum	0.698864	1.22191	-0.26818	15	2	0.002	Y
	SAFe	0.793726	1.109201	-0.21267	7	2	0.18	N
	Other	0.716008	1.014296	-0.15089	3	2	1	N

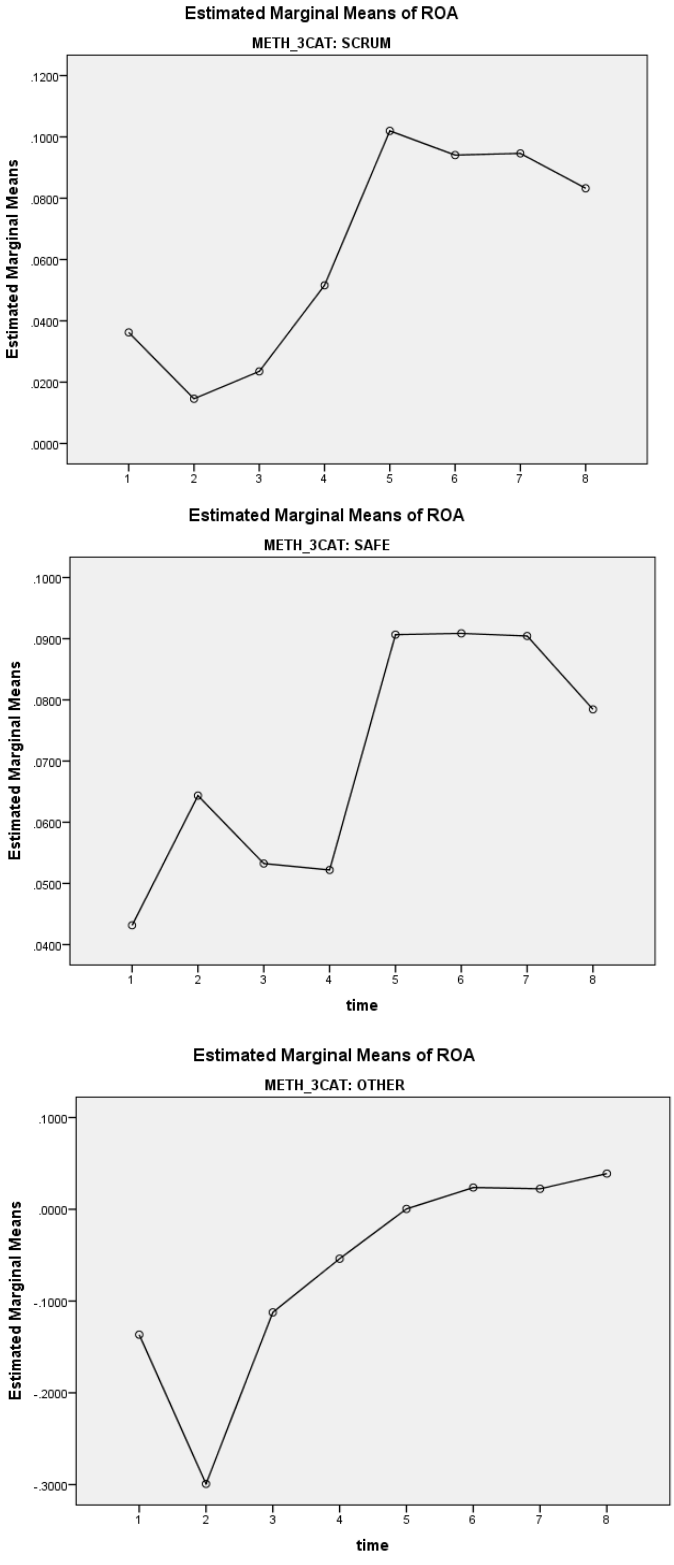


Figure 4-5 MAIN EFFECTS of ROA by Agile Method

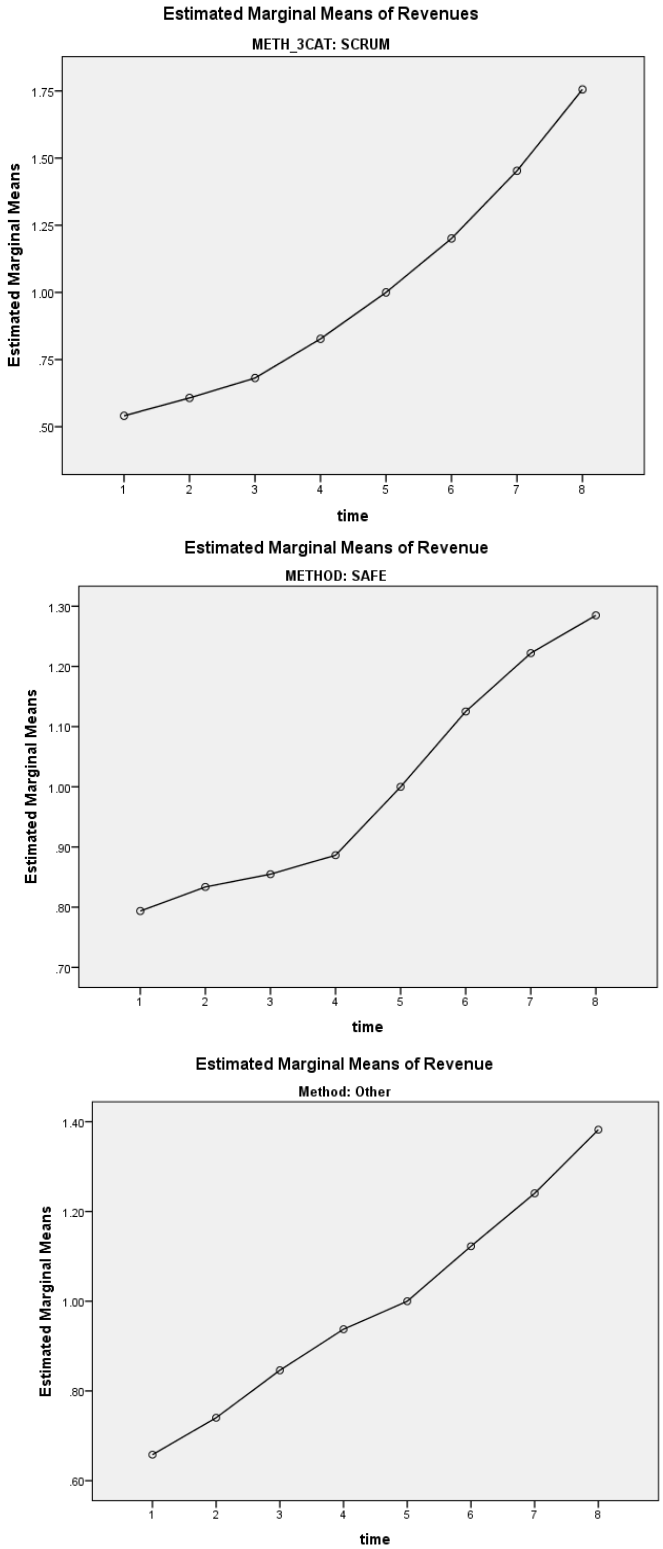


Figure 4-6 Main effects of Revenue by Agile Method

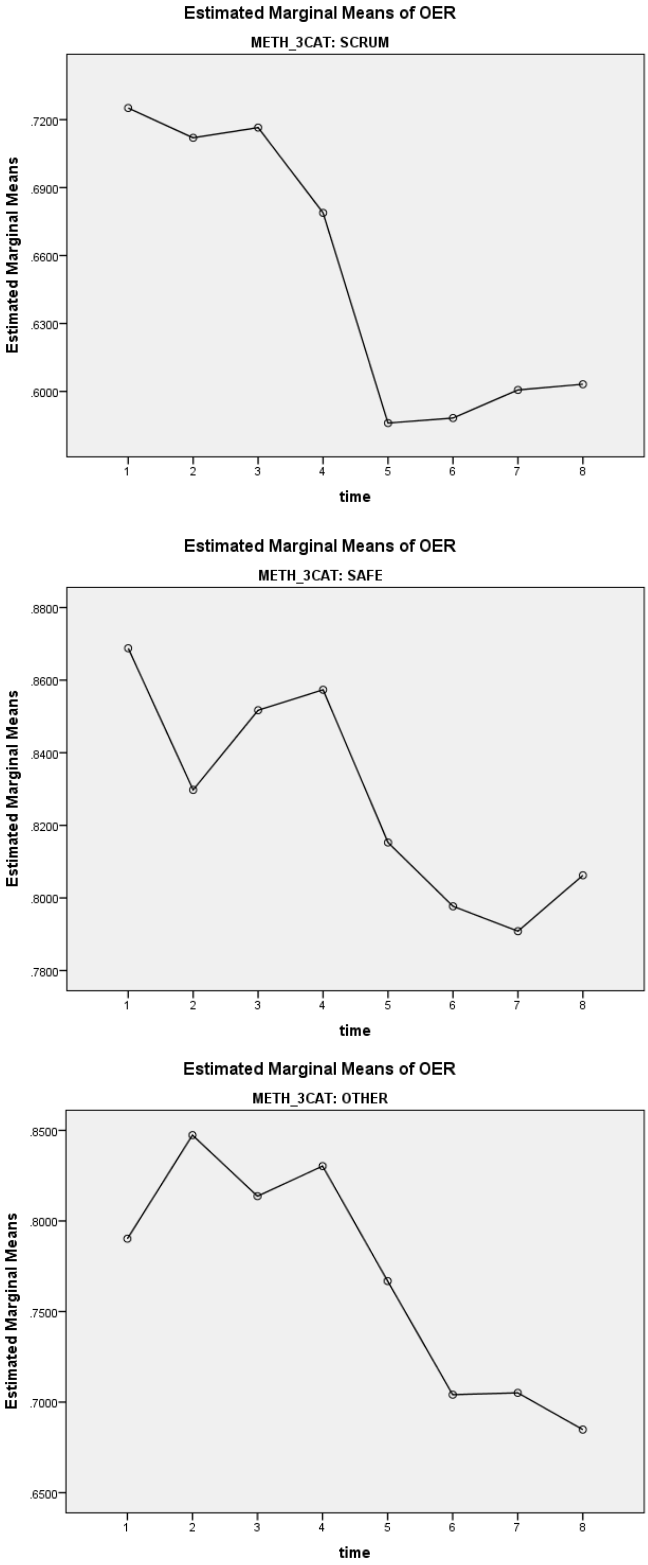


Figure 4-7 Main effects OF ROE by method

Chapter 5: Discussion of Conclusions

5.1 Conclusions

This study investigated the organization level performance impact of switching to the use of Agile frameworks. Organizations that shifted to Agile methods showed a reduction (improvement) on OER and an increase in ROA. While Revenues also increased after the implementation of Agile methods, the change in Revenue cannot be attributed to the intervention and is likely due to normal revenue growth. As such, only the first and fifth hypotheses are supported.

The study was not able to show a statistically significant difference in performance based on which framework was utilized. The Sign test performed on each dependent variable for Scrum, SAFe, and Other indicated a substantially higher median change in Scrum than SAFe, while the same test with ROA indicated a higher median improvement in SAFe than Scrum. Qualitatively, the main effects plots show similar behavior to the combined data for all three variables. Both Scrum and SAFe seemed to perform better than other methods in both OER and Revenues, though Scrum showed the smallest median difference in ROA after it was implemented. That said, for ROA both Scrum and SAFe show a statistically significant improvement. For OER and Revenues, only Scrum showed a significant improvement. Interestingly enough, for ROA, SAFe showed a much higher median difference than scrum for ROA.

5.2 Discussion

Agile methods seek to increase the value delivered through the business through better prioritization and collaboration. They also seek to drive increased customer satisfaction. It is believed that these factors should result in accelerated growth in

Revenues, and the expected result was an increase in the rate of growth of Revenues, likely with a time lag. This study could not identify such a change. While it is possible that the lag is greater than the 3-4 year time frame after Agile methods are introduced, at this point there is no evidence that this is the case.

Operationally, the improvement of organizations is as expected. Typically, operating expenses scale in conjunction with Revenues, as organizations rely on additional resources to respond to increasing demand. An increase in operational efficiency would allow an organization to increase Revenues without a corresponding increase in costs, or allow them to maintain similar levels of Revenues while cutting existing costs. Because the easiest improvements are often implemented first, later improvements would likely result in small enough improvements that they would be difficult to identify in an organizational level study, thus a relatively stepwise reduction in OER is logical.

ROA will increase with higher Revenues, but will also increase with lower OER. As operating costs are reduced, net profit will increase. Likewise, Revenue increases are likely to result in increased ROA. As such, the expected performance would be both a stepwise improvement due to reduced operating expenses with subsequent increase in profitability over time due to ongoing top line growth. Instead, performance mirrors that of OER, again indicating a lack of top line growth attributable to the implementation of Agile methods.

Because the improvements are only in relation to ROA and OER, the advantage must come from overall efficiency of operations, as top line growth should result in stronger Revenue performance.

Initially, one might expect Scrum to show more improvement than SAFe across the board, as many Agile experts consider SAFe more restrictive, limiting, and less flexible. That said, because Scrum recommends a very flat organization and SAFe adds multiple roles at the program and portfolio level, it is logical to assume that Scrum carries less overhead than SAFe, even for organizations of roughly the same size. It may be that SAFe, with additional personnel in product management roles can more effectively meet the needs of the customer base and more effectively identify strategic initiatives and take advantage of them, leading to higher profitability, even if they aren't quite as efficient as Scrum operationally.

In retrospect, the inability of this study to differentiate between Agile frameworks is not surprising. While there is significant criticism of more restrictive methodologies like SAFe that a lack of degrees of freedom will lead to lower impact of transition, the fact that the more typical operational data shows a moderate improvement for Scrum despite 'properly coached' team performance showing productivity and quality improvements of up to an order of magnitude indicates that the actual implementations are generally not as impressive. Without significant differences in performance, distinguishing performance between similar models will be difficult (Sutherland, 2014).

Additionally, the granularity of this study may not be sufficient to derive any difference between methodologies, or any differential that exists may be so small as to be very difficult to detect. Of the methodologies identified, many methods had only one firm identified with adequate data, so a statistical significance could not be shown. When Scrum and SAFe were compared to 'other', it is possible that some methods are better, some are worse, and differences are cancelled out.

In short, it is likely that the actual implementations of agile frameworks vary to such a degree in terms of technical practice, team empowerment and dynamics, and product alignment that the advantages of one framework over another are a smaller factor than how 'Agile' a given organization is becoming.

5.3 Contribution to the Body of Knowledge

The results of this research have practical applications across multiple fields. Firms are likely to operate more efficiently and effectively when using Agile frameworks instead of traditional project management approaches. This study implies a significant role in organization operating methodology and can provide impetus for organizational change.

For firms that are currently using Agile methods, this study may provide direction as to where continuous improvement efforts may be concentrated. Because efficiency appears to be the primary benefit at this time, it is possible that a leaner operating model is the result of increased productivity, favoring reduction of resources and costs associated instead of increased overall throughput or productivity.

Improved alignment with Business Priorities, improved quality, and greater productivity should drive top line revenue performance and does not appear to. It is likely that the implementations of Agile methods are addressing only operational concerns and not adequately addressing the business and development alignment or effectively prioritizing highest value work and limiting organizational work in progress. From a management perspective, improvement in these areas is paramount, especially in organizations that are already utilizing Agile methods.

Additionally, this study provides a novel way to address operational research. This is the first study of this type to utilize Change Point Analysis, and the first study in any field to use Change Point Analysis to identify a change point and verified using the Chow test. As such, it provides a framework for future operational and engineering research to effectively use longitudinal data.

5.4 Future Research

Instead of focus on specific frameworks, there are multiple tools that seek to gauge a level of overall Agility. For example, Mike Cohn's comparative agility survey creates a World Agility Index based on multiple factors, so you can see how 'agile' an organization or a team is relative to other organizations or teams (Cohn, n.d.). Other sites offer different measures such as agility health (Agility Health, 2017). Correlation between level of agility and bottom line performance would potentially offer more management insight into operating models, and may allow for differentiation based on Agile framework used. This may also offer a point of comparison between frameworks.

Additionally, tools like Comparative Agility rank organizations based on a variety of categories, which would allow comparison of Revenue performance with practices that should lead to greater alignment with business priorities and drivers of customer satisfaction (Cohn, n.d.).

References

Aba, E. K. (2016). Impact of ISO 9001 certification on firms financial operating performance. *International Journal of Quality and Reliability Management, 33*(1), 78-89.

Accardi-Petersen, M. (2011). *Agile Marketing.* New York: Apress.

Adam, E. J. (1994). Alternative quality improvement practices and organizational performance. *Journal of Operations Management*, 27-44.

Adkins, L. (2014, Oct 6). *Tha Agile Coaches' Coach Shares Her View on SAFe*. Retrieved Sept 19, 2017, from InfoQ articles: https://www.infoq.com/articles/agile-coaches-coach-view-safe

Agile Alliance. (n.d.). *Scrum of Scrums*. Retrieved Aug 28, 2017, from Agile Alliance Glossary: https://www.agilealliance.org/glossary/scrum-of-scrums/#q=~(filters~(postType~(~'page~'post~'aa_book~'aa_event_session~'aa_experience_report~'aa_glossary~'aa_research_paper~'aa_video)~tags~(~'scrum*20of*20scrums))~searchTerm~'~sort~false~sortDirection~'as

Agility Health. (2017). *Agility Health Radars*. Retrieved April 28, 2017, from Agility Health: https://agilityhealthradar.com/radars/

Al-Baik, O. a. (2015). The kanban approach, between agility and leanness: a systematic review. *Empirical Software Engineering, 20*(6), 1861-1897.

Al-Baik, O. M. (2015). the kanban approach, between agility and leanness: a systematic review. *Empirical Software Engineering*, 1861-1897.

Albers, W. B. (2000). Size and power of pretest procedures. *Annals of Statistics, 28*, 195-214.

Amaral, C. e. (2015). Early postnatal nociceptive stimulation results in deficits of spatial memory in male rats. *Neurobiology of Learning and Memory, 125*, 120-125.

Ambler, S. (2008). *Agile adoption survey.* Retrieved March 15, 2017, from Ambisoft: www.ambisoft.com

Ambler, S. (2014, April 10). *Extending the Agile Manifesto*. Retrieved Sept 21, 2017, from Disciplined Agile Delivery: http://www.disciplinedagiledelivery.com/extending-the-agile-manifesto/

Ambler, S. (n.d.). *The Disciplined Agile Manifesto*. Retrieved Sept 23, 2017, from Disciplined Agile Delivery: http://www.disciplinedagiledelivery.com/disciplinedagilemanifesto/

Anderson, D. J. (2010). *Kanban: Successful Evolutionary Change for your Technology Business.* Blue Hole Press.

Anderson, J. R. (1995). A path analytic model of a theory of quality management underlying the Deming management method: preliminary empirical findings. *Decision Sciences, 26*(5), 637-658.

Armel. (2012, January 23). *Top Performing Projects Use Small Teams.* Retrieved Sept 3, 2017, from Quantitative Software Management: http://www.qsm.com/blog/2012/top-performing-projects-use-small-teams

Association of Modern Technology Professionals. (n.d.). *Project Management Body of Knowledge (PMBOK) Guide*. Retrieved Sept 18, 2017, from IT Knowledge Portal: http://www.itinfo.am/eng/project-management-body-of-knowledge-pmbok-guide/

Bapuji, H. D. (2011). Connecting external knowledge usage and firm performance: an empirical analysis. *Journal of Engineering and Technology Management, 28*, 215-231.

Barkaui, K. (2014). Quantitative Approaches for Analyzing Longitudinal Data in Second Language Research. *Annual review of applied linguistics*, 65-101.

Beck, K. B. (2001). *AgileManifesto.org.* Retrieved from The Manifesto for Agile Software Development.

Beck, K. e. (2001). *The Agile Manifesto*. Retrieved Jan 3, 2017, from Agile Alliance: https://www.agilealliance.org/agile101/the-agile-manifesto/

Bennett, A. L. (2014, June). Certified Scrum Master Training Deck.

Black, N. (2014, Oct 14). *A Brief History of Time(lines): Henry Gantt and his Revolutionary Chart*. Retrieved Sept 21, 2017, from OnePager Community Blog: https://www.onepager.com/community/blog/a-brief-history-of-the-gantt-chart/

BMC Software Inc Form 10K for fiscal year ended March 31, 2010. (2010, May). Retrieved from SEC Archives EDGAR: https://www.sec.gov/Archives/edgar/data/835729/000119312510112656/d10k.htm

BMC Software Inc. Form 10-K Year Ended March 31, 2003. (n.d.). Retrieved from SEC Archives EDGAR.

Bogle, J. C. (2008). *Enough: True Measures of Money, Business, and Life.* New York: Wiley.

Bureau of Labor Statistics. (2017, Nov 13). *CPI - All Urban Consumers*. Retrieved from Databases, Tables, and Calculators by Subject: https://data.bls.gov/timeseries/CUUR0000SA0L1E?output_view=pct_12mths

Campbell, K. a.-V. (2008). Gender Diversity in the Boardroom and Firm Financial Performance. *Journal of Business Ethics, 83*(3), 435-451.

Cardozo, E. e. (2010). SCRUM and Productivity in Software Projects: A Systematic Review. *EASE*.

Cassidy, M. M. (2002). Movement Related Changes in Cynchronization in Human Basal Ganglia. *Brain*, 1235-1246.

Casson, R. J. (2014). Understanding and checking the assumptions of linear regression: a primer for medical researchers. *Clinical and Experimental Ophthalmology, 42*(6), 590-596.

Cervone, H. (2011). Understanding Agile Project Management Methods using Scrum. *OCLC Systems and Services , 27*(1), 18-22.

Cervone, H. F. (2011). Understanding agile project management methods using Scrum. *OCLC Systems and Services, 27*(2), 18-22.

Chambers, J. C. (1983). *Graphical methods for data analysis.* Pacific Grove: Wadsworth and Brooks.

Chang, Y. B. (2011). Does RFID improve firms' financial performance? an empirical analysis. *Information Technology and Management, 12*(3), 273-285. Retrieved Jan 2, 2017, from http://proxygw.wrlc.org/login?url=http://search.proquest.com/docview/881408424?accountid=11243

Clark, W. (1922). *The Gantt Chart a working tool of Management.* New York: Ronald Press Company.

Coelho, E. a. (2012). Effort Estimation in Agile Software Development using Story Points. *Foundation of Computer Science, 3*(7), 7-10.

Cohn, M. (2009). *Succeeding with Agile.* Upper Saddle River: Addison Wesley.

Cohn, M. (n.d.). *Comparative Agility*. Retrieved March 15, 2017, from https://www.comparativeagility.com

Conforto, E. C. (2016). Agile project management and stage-gate model- A hybrid framework for technologyt-based companies. *Journal of Engineering and Technology Management, 40*, 1-16.

Corrigan, J. (1994). Is ISO 9000 the path to TQM. *Quality Progress, 27*(5), 33-36.

Cram, P. F. (2003). The Impact of a Celebrity Promotional Campaign on the use of Colon Cancer Screening: The Katie Couric Effect. *Internal Medicint, 163*, 1601-1605.

CSG International Inc 2010 Form 10-K Annual Report. (2011, Feb). Retrieved from Investor Relations of CSG: http://ir.csgi.com/secfiling.cfm?filingID=1193125-11-59556&CIK=1005757

David. (n.d.). *The Power Advantage of Within Subjects Designs*. Retrieved Oct 13, 2017, from Statistics Solutions: http://www.statisticssolutions.com/the-power-advantage-of-within-subjects-designs/

Decker, S. (2003). *Yahoo Inc. Form 10-K For the fiscal year 2002.* Yahoo.

Decker, S. (2005). *Form 10-K for the fiscal year 2004.* Yahoo.

Deemer, P. B. (2010). *Scrum Primer version 1.2.* Retrieved July 15, 2017, from www.ScrumPrimer.com: http://goodagile.com/scrumprimer/scrumprimer.pdf

Deemer, P. B. (2012). *Scrum Primer 2.0 A lightweight guide to the theory and practice of scrum.* Retrieved Sept 13, 2017, from Scrum Primer: www.scrumprimer.com

Dinov, I. (2012). *F Distribution Tables*. Retrieved Dec 1, 2017, from Statistics Online Computational Resource (SOCR): http://www.socr.ucla.edu/applets.dir/f_table.html

Disciplined Agile Consortium. (n.d.). *Disciplined Agile 2.X a Process Decision Framework*. Retrieved May 1, 2017, from Disciplined Agile: http://www.disciplinedagiledelivery.com/

Downey, S. a. (2013). Scrum Metrics for Hyperproductive Teams: How they Fly like Fighter Aircraft. *47th Hawaii International Conference on System Sciences*, (pp. 4870-4878). Wailea.

Duarte, A. B. (2011). Operational Practices and Financial Performance: an Empirical Analysis of Brazilian Manufacturing Companies. *Brazilian Administration Review, 8*(4), 395-411.

Duran-Encalada, J. a.-R. (2015). Effects of family ownership, debt, and board composition on Mexican Firms Performance. *International Journal of Financial Studies*.

Dyba, T. D. (2009). What do we know about Agile Software Development. *IEEE Software 2009, 26*(5), 6-9.

Ebert, C. a. (2017). Scaling Agile. *IEE Software, 34*(6), 98-103.

Emergn. (2016). *Value Flow Quality: Why Change.* Boston: Emergn.

Ernst, A. A. (2017). Regression assumptions in clinical psychology research practice- a systematic review of common misconceptions. *Peer J, 5*.

Ewell, J. (2011). *Who is doing Agile marketing?* Retrieved Nov 28, 2017, from Agile Marketing: http://www.agilemarketing.net/whos-doing-agile-marketing/

Fernandez, D. J. (2009). Agile Project Management - Agilism versus Traditional Approaches. *The Journal of Computer Information Systems, 49*(2), 10-17.

Field, A. (2011). *Discovering Statistics Using SPSS* (3rd ed.). Thousand Oaks, Ca: SAGE Publications.

Foffani, G. P. (2003). 300 Hz Subthalamic Oscillations in Parkinson's Disease. *Brain*, 2153-2163.

Ford, H. (1922). *My Life and Work.* Garden City, NY: Doubleday, Page, & Co.

Fornell, C. M. (2006). Customer Satisfaction and stock prices: High returns at low risk. *Journal of Marketing, 70*(1), 3-14.

Forte, T. (2016, Oct 1). *Theory of Constraints 102: The Illusion of Local Optima.* Retrieved from Praxis: https://praxis.fortelabs.co/theory-of-constraints-102-local-optima-3ca8d348f146

Foster, S. J. (2007). Does Six Sigma improve performance? *Quality Management Journal*, 7-20.

Frost, J. (2015, Sept 17). *Repeated Measures Designs: Benefits, Challenges, and an ANOVA Example.* Retrieved March 2, 2017, from The Minitab Blog: http://blog.minitab.com/blog/adventures-in-statistics-2/repeated-measures-designs-benefits-challenges-and-an-anova-example

Gavil, P. T. (2009, Aug). Use of Change-Point Analysis for Process Monitoring and Control. A better method for trend analysis than CUSUM and control charts. *BioPharm International, 22*(8).

George, S. W. (1998). *Total Quality Management: Strategies and Techniques Proven at today's most successful companies.* New York: Wiley.

Ghani, I. (2015, Oct 31). A survey based analysis of agile adoption on performances of IT Organizations. *Journal of Korean Society for Internet Information, 16*(5), 87-92.

Grace-Martin, K. (n.d.). *Approaches to Repeated Measures Data: Repeated Measures ANOVA, Marginal, and Mixed Models*. Retrieved Oct 1, 2017, from The Analysis Factor: http://www.theanalysisfactor.com/repeated-measures-approaches/

Hall, R. (n.d.). *Within Subjects Designs*. Retrieved Aug 14, 2017, from Psychology World: https://web.mst.edu/~psyworld/within_subjects.htm

Hannon, J. (2014, February 7). *Yes Scrum Does Work in Education.* Retrieved Dec 13, 2016, from Scrum Alliance: https://www.scrumalliance.org/community/articles/2014/february/yes-scrum-does-work-in-education

Hansen, G. W. (1989). Determinants of Firm Performance: The Relative Importance of Economic and Organizational Factors. *Strategic Management Journal, 10*(5), 399-411.

Harrison, J. B. (n.d.). *Taking the Mystique out of Kanban Systems*. Retrieved Oct 1, 2017, from Hands on Group: http://www.handsongroup.com/lean-articles/taking-the-mystique-out-of-kanban-systems/

Hayes, W. (2017, February 20). *Five Perspectives on Scaling Agile.* Retrieved from Software Engineering Institute Carnegie Mellon University:

https://insights.sei.cmu.edu/sei_blog/2017/02/five-perspectives-on-scaling-agile.html

Heishman, A. (2015). Effectiveness of computerized working memory training on math achievement and other transfer effects in children with ADHD and math difficulties. *George Washington University*.

Hendricks, K. S. (1997). Does implementing an effective TQM program actually improve operating performance? Empirical evidence from firms that have won quality awards. *Management Science*, 1258-1274.

Heredia, A. G.-G.-S.-D. (2014). Agile practices adapted to mass market application development. *Software: Evolution and Process, 26*(9), 818-828.

Hermalin, B. a. (1991). The Effects of Board Composition and Direct Incentives on Firm Performance. *Financial Management*.

Hietschold, N. (2014). Measuring critical success factors of TQM implementation succesfully- a systematic literature review. *International journal of production research, 52*(21), 6254-6272.

Highsmith, J. (2001). *A History of the Agile Manifesto.* Retrieved from AgileManifesto.org: http://agilemanifesto.org/history.html

Hoegstron, F. a. (2017, Nov 27). US People Strategy and Innovation:Agile Transformation. Bentonville, Arkansas.

Howitt, D. &. (2011). *Introduction to Research Methods in Psychology.* (3rd ed.). Harlow, Essex: Pearson Education Limited.

Hughey, D. (2009). *Comparing Traditional Systems Analysis and Design with Agile Methodologies*. Retrieved Aug 21, 2017, from University of Missouri, St. Louis Information Systems: http://www.umsl.edu/~hugheyd/is6840/waterfall.html

Hung, C. S. (2012). An empirical study of the relationship between a self service technology investment and firm financial performance. *Journal of Engineering and Technology Management, 29*, 62-70.

Hung, C. S. (2012). An Empirical Study of the relationship between a self service technology investment and firm financial performance.". *Journal of Engineering and Technology Management, 29*, 62-70.

Hwang, D. M. (2015). Mediating effect of IT enabled capabilitis on competitive performance outcomes: an empirical investigation of ERP implementation. *Journal of Engineering and Technology Management, 36*, 1-23.

Inc. (2015). *Leandog.* Retrieved Nov 21, 2017, from Inc. 5000: https://www.inc.com/profile/leandog

Investing Answers. (n.d.). *Opeerating Expense Ratio*. Retrieved Sept 21, 2017, from InvestingAnswers: www.investinganswers.com/final-statement-analysis/operating-expense-ratio-oer-2800

Investing Answers. (n.d.). *Operating Expense Ratio (OER)*. Retrieved Aug 21, 2017, from Investing Answers: www.investinganswers.com/final-statement-analysis/return-equity-roe-916

Italtel. (2011, April 6). *Italtel Financial Reports.* Retrieved from Italtel Group 2010 Annual Report: http://www.italtel.com/content/uploads/2016/01/Italtel-Group-Annual-Report-2010-eng.pdf

Italtel. (2016, June 10). *Italtel SPA Group 2015 Directors' Report and Consolidated Financial Statements.* Retrieved from Italtel Financial Statements: http://www.italtel.com/content/uploads/2016/01/Italtel_SpA_2015_Consolidated_financial_statements.pdf

Italtel Group 2011 Director's Report and Consolidated Financial Statements. (2012, Dec 21). Retrieved from Italtel Financial Reports: http://www.italtel.com/content/uploads/2016/01/Italtel-Group-Consolidated-Financial-Statements-2011-eng.pdf

Italtel Group 2013 Directors' Report and Consolidated Financial Statements. (2014, March 28). Retrieved from Italtel Financial Reports: http://www.italtel.com/content/uploads/2016/01/Italtel-GroupItaltel-Annual-Report-2013.pdf

Ittner, C. a. (2003, Nov). Coming up short on nonfinancial performance measurement. *Harvard Business Review*.

Johnsen, A. (2012). Quantifying the effect of using kanban versus scrum: a case study. *IEEE Software, 29*(5), 47-53.

Joyce, W. N. (2003). *What Really Works: The 4+2 Formula for Sustained Business Success.* New York: Harper Business.

Kaplan, R. a. (1992). The balanced scorecard: Measures that drive performance. *Harvard Business Review, 70*(1), 71-79.

Karekar, H. (2016). *Scaled Agile Framework.* Retrieved from Case Study: Amdocs: http://scaledagileframework.com/amdocs-case-study

Kautz, K. J. (2014). The perceived impact of the Agile Development and Project Management Method Scrum on Information Systems and Software Development Productivity. *Australasian Journal of Information Systems, 18*(3).

Kessel, C. (2013, January 1). *Softare History: Waterfall, the process that wasn't meant to be.* Retrieved August 23, 2017, from OBS Global Blog: https://info.obsglobal.com/blog/2013/01/software-history-waterfall-the-process-that-wasnt-meant-to-be

Kesselman, H. R. (1980). Testing the validity conditions of repeated measures F tests. *Psychological Bulletin, 87*, 479-481.

KIDASA Software. (n.d.). *Henry Gantt's Legacy is the Gantt Chart*. Retrieved Sept 21, 2017, from Gantchart.com: http://www.ganttchart.com/history.html

Laerd Statistics. (2015). *Simple linear regression using SPSS Statistics.* Retrieved from Statistical tutorials and software guides.: https://statistics.laerd.com/premium/spss/lr/linear-regression-in-spss-22.php

Laerd Statistics. (2015). *Three-way repeated measures ANOVA using SPSS Statistics.* Retrieved March 2, 2017, from Statistical tutorials and software guides: https://statistics.laerd.com/premium/spss/ftwrma/three-way-repeated-measures-anova-in-spss-5.php

Laerd Statistics. (2015). *Three-way repeated measures ANOVA using SPSS Statistics.* Retrieved March 15, 2017, from Statistical tutorials and software guides.: https://statistics.laerd.com/premium/spss/ftwrma/three-way-repeated-measures-anova-in-spss-22.php

Laerd Statistics. (2015). *Wilcoxon signed-rank test using SPSS Statistics.* Retrieved July 03, 2017, from Statistical tutorials and software guides.: https://statistics.laerd.com/premium/spss/wsrt/wilcoxon-signed-rank-test-in-spss-25.php

Laerd Statistics. (n.d.). *Paired-samples t test using SPSS Statistics.* Retrieved June 30, 2017, from Statistical Tutorials and software guides: https://statistics.laerd.com/

Lakshmi Tulasi, C. R. (2005). Review on Theory of Constraints. *International Journal of Advances in Engineering and Technology, 3*(1).

Lane, D. (n.d.). *Advantages of Within-Subjects Designs*. Retrieved Sept 24, 2017, from Hyperstat online: http://davidmlane.com/hyperstat/within-subjects.html

Larman, C. a. (n.d.). *Large Scale Scrum*. Retrieved May 1, 2017, from Large Scale Scrum: https://less.works/

Lean Lab. (n.d.). *Why and What is Kanban*. Retrieved Sept 25, 2017, from Lean Lab: http://www.leanlab.name/why-and-what-is-kanban

Lean Manufacturing Tools. (2017). *Kanban*. Retrieved Sept 25, 2017, from Lean Manufacturing Tools: http://leanmanufacturingtools.org/kanban/

LeanKit. (n.d.). *What is Kanban?* Retrieved Aug 29, 2017, from Leankit: https://leankit.com/learn/kanban/what-is-kanban/

Lee, G. a. (2010, March). Toward Agile: an Integrated Analysis of Quatitative and Qualitative Field Data on Software Development Agility. *MIS Quarterly, 34*(1), 87-114.

Leftingwell, D. (2017). *About SAFe*. Retrieved Sept 19, 2017, from the Scaled Agile Framework: http://www.scaledagileframework.com/about/

Leftingwell, D. e. (n.d.). *SAFe*. Retrieved May 1, 2017, from Scaled Agile Framework: http://www.scaledagileframework.com/

Lima, M. R. (2000). Quality certification and performance of Brazilian firms: an empirical study. *International Journal of Production Economics, 66*(2), 143-147.

Linders, B. (2013, August 21). *Scrum for Education - Experiences from eduScrum and Blueprint Education.* Retrieved Dec 13, 2016, from InfoQ : https://www.infoq.com/articles/scrum-education

Longin, M. a. (2015, June 8). *Lean Kanban North America 2015.* Retrieved from Ultimate Software: Moving to a Data Driven Approach: http://schd.ws/hosted_files/lkna15/09/LKNA%20-%20Ultimate%20Software%20Moving%20to%20a%20Data%20Driven%20Approach%20-%20Final%20%281%29.pdf

March, J. S. (1997). Organizational Performance as a Dependent Variable. *Organizational Science, 8*(6), 698-706.

Maxwell, S. a. (2004). *Designing experiments and analyzing data: A model comparison perspective.* NY, NY: Psychology Press.

May, J. Y. (2016). Play Ball: Bringing Scrum into the Classroom. *Journal of information Systems Education, 27*(2), 87-92.

McGuire, J. B. (1990). Perceptions of Firm Quality: A Cause or Result of Firm Performance. *Journal of Management, 16*(1), 167-180.

Merrit, C. (n.d.). *The Size Limits for Small-Cap, Mid-Cap & Large-Cap Stocks.* Retrieved Dec 24, 2016, from Zacks: www.finance.zacks.com/size-limits-smallcap-midcap-largecap-stocks-5895.html

Miller, D. a. (2002). Spotting Management Fads. *Harvard Business Review*.

Mir, M. M. (2016). The impact of standardized innovation management systems on innovation capability and business performance: An empirical study. *Journal of Engineering and Technology Management, 41*, 26-44.

Misangyi, V. L. (2006, Jan). The Adequacy of Repeated Measures Regression for Multilevel Research : Comparisons with Repeated Measures ANOVA, Multivariate Repeated Measures ANOVA, and Multilevel Modeling across Various Multilevel Research Designs. *Organizational Research Methods*, 5-28.

Morris, B. (2006, July 11). New Rule: Look out, not in. *Fortune*.

Mowery, B. D. (2011, Dec). The Paired t-Test. *Pediatric Nursing, 37*(6).

Murphy, K. R. (2016). Mend it or End it redirecting the search for interactions in the organizational sciences. *Organizational Research Methods*, *20*(4), 549-573.

Mustonen, A.-M. L. (2012). Application of change-point analysis to determine winter sleep patterns of the racoon dog from body temperature recordings and a multi-faceted dietary and behavioral study of wintering. *BMC Ecology, 12*(27).

Neumarker, N. (2017, Nov 31). SVP Software Development, Verscend. (A. Bennett, Interviewer)

New, S. (2007). Celebrating the enigma: the continuing puzzle of the Toyota Production System. *International Journal of Production Research, 45*(16), 3545-3554.

Ongore, V. P. (2015). Board composition and financial performance:empirical analysis of companies listed at the Nairobi securities exchange. *International Journal of Economics and Financial Issues*(5), 23-43.

Owen, D. (2011, Dec 1). *The Advantages and Disadvantages of Repeated Measures*. Retrieved May 19, 2017, from Bangor University Blogging: https://dsowen.wordpress.com/2011/12/01/the-advantages-and-disadvantages-of-repeated-measures/

Paypal 2015 Annual Report. (2016, May). Retrieved from Paypal Investor Relations.

PayPal. (2015, Sept 18). *Paypal Enterprise Transformation.* Retrieved from Paypalobjects.com: https://www.paypalobjects.com/webstatic/en_US/mktg/pages/stories/pdf/paypal_transformation_whitepaper_sept_18_2015.pdf

Pomar, F. A.-M.-C. (2014). Understanding Sprint Velocity fluctuations for improved project plans with Scrum: a case study. *Journal of Software: Evolution and Process, 26*(9), 776-783.

Prieto, F. J. (2016, August 31). *The Agile Classroom: Embracing an Agile Mindset in Education.* Retrieved Nov 29, 2017, from Labratoria: https://medium.com/laboratoria/the-agile-classroom-embracing-an-agile-mindset-in-education-ae0f19e801f3

Przasnyski, Z. a. (2002). Stock performance of malcolm baldrige national quality award winning companies. *Total Quality Management, 13*(4), 475-488.

Quantatative Software Management Associates(QSMA). (2008). *The Agile Impact Report: Proven Performance Metrics from the Agile Enterprise.* QSMA Inc. Retrieved from Quantatative Software Management Associates Inc.: http://qsma.com/books-reports/

Rabon, B. M. (2015, June 19). *Scaling Scrum: a brief comparison of DaD, LeSS, and SAFe*. Retrieved Sept 21, 2017, from Linkedin Pulse: https://www.linkedin.com/pulse/scaling-scrum-brief-comparison-dad-less-safe-brian-m-rabon-cst-pmp/

Rico, D. (2007). *Effects of Agile Methods on Website Quality for Electronic Commerce.* University of Maryland University College.

Rico, D. (2008). What is the ROI of Agile vs. Traditional Methods? An analysis of XP, TDD, Pair Programming, and Scrum (Using Real Options). *TickIT International, 10*(4), 9-18.

Rico, D. H. (2009). *The Business Value of Agile Software Methods.* India: Cengage Learning.

Rico, D., Sayani, H., & Sone, S. (2009). *The Business Value of Agile Software Methods.* Stamford, Ct: Cengage Learning.

Rigby, D. S. (2016, April 20). *The Secret History of Agile Innovation.* Retrieved August 1, 2017, from Harvard Business Review: https://hbr.org/2016/04/the-secret-history-of-agile-innovation

Rostami, S. R. (2016). The Effect of Corporate Governance Components on Return on Assets and Stock Return of Companies Listed in Tehran Stock Exchange. *Procedia Economics and Finance, 36*, 137-146.

Royce, W. (1970). Managing the Development of Large Software Systems. *Proceedings of IEEE WESCON 26*, (pp. 1-9).

Rubin, K. (2013). *Essential Scrum.* Ann Arbor, MI: Pearson Education.

Salesforce 2012 Annual Report. (2013). Retrieved from Salesforce Investors: http://s1.q4cdn.com/454432842/files/doc_financials/2012/fy12_annual_report.pdf

Salesforce.com 2008 Annual Report. (2009). Retrieved from Salesforce Investors: http://s1.q4cdn.com/454432842/files/doc_financials/2008/fy08_annual_report.pdf

Salesforce.com. (2010). *Transforming your organization to Agile the inside story of saleforce.com's transformation from waterfall to agile.* San Francisco: Salesforce.com.

Scaled Agile. (2015, Feb 11). Leading SAFe . *Leading SAFe Facilitators guide*.

Scaled Agile. (2017, June 16). *Core Values.* Retrieved Sept 19, 2017, from Scaled Agile Framework: http://www.scaledagileframework.com/safe-core-values/

Scaled Agile. (2017, June 2). *Program Increment.* Retrieved Sept 19, 2017, from Scaled Agile Framework: http://www.scaledagileframework.com/pi-planning/

Scaled Agile. (2017, April 2). *Program Level.* Retrieved Sept 19, 2017, from Scaled Agile Framework: http://www.scaledagileframework.com/program-level/

Scaled Agile. (n.d.). *Permissions FAQ*. Retrieved Sept 18, 2017, from Scaled Agile Framework: https://www.scaledagile.com/about/about-us/permissions-faq/

Scaled Agile. (n.d.). *The Scaled Agile Framework*. Retrieved Sept 21, 2017, from Scaled Agile Framework: http://www.scaledagileframework.com/

Schaller, J. (2005). "Kanban - Do it now but do it right" Workshop Illustrates the Importance of Kanban as a Tool in Lean Production. *Association of Mechanical Engineers Target, 21*(2), 43-50.

Schwaber, K. (1997). The Scrum Development Process. *Business Object Design and Implementation* (pp. 117-134). London: Springer.

Schwaber, K. (2013, Aug 6). *UnSAFe at any speed*. Retrieved Sept 19, 2017, from Ken Schwaber Blog: https://kenschwaber.wordpress.com/2013/08/06/unsafe-at-any-speed/

Sedge, t. (2014, July 15). *In defence of the Scaled Agile Framework (SAFe)*. Retrieved Sept 19, 2017, from The Ambitious Manager: http://www.ambitiousmanager.com/defence-scaled-agile-framework-safe/

Seltman, H. J. (2015). *Experimental Design and Analysis.* http://www.stat.cmu.edu/~hseltman/309/Book/.

Sferlazza, F. (2011). *LeSS Adoption at Italtel.* Retrieved from LeSS Case Studies: https://less.works/case-studies/italtel.html

Shafer, S. M. (2012). The effects of Six Sigma on corporate performance: An empirical investigation. *Journal of Operations Management*, 521-532.

Shalloway, A. (2011). Demystifying Kanban. *Cutter IT Journal, 24*(3), 12-17.

Singh, V. R. (2013). Analysis of repeated measurement data in the clinical trials. *Journal of Ayurveda and Integrative medicine, 4*(2), 77-81.

Sliger, M. (2008). Agile Project Management and the PMBOK guide. *PMI Global Congress 2008.* Denver, Co: Project Management Institute.

Soh, P. (2017, June 2). *Accenture Acquires SolutionsIQ, Adds Leading Agile Transformation Expertise and Services.* Retrieved Nov 20, 2017, from Accenture Newsroom: https://newsroom.accenture.com/news/accenture-acquires-solutionsiq-adds-leading-agile-transformation-expertise-and-services.htm

Srinivasasan, R. (2016, April 22). *Large Scale Scrum, More with LeSS.* Retrieved Sept 21, 2017, from Slideshare: https://www.slideshare.net/ramvasan/large-scale-scrum-more-with-less

Stavru, S. (2014). A critical examination of recent industrial surveys on agile method usage. *Journal of Systems and Software., 94*, 87-97.

Stoica, M. G.-M. (2016). Analyzing Agile Development- from Waterfall Style to Scrumban. *Informatica Economica*, 5-14.

Suetin, S. V. (2016). Results of agile project management implementation in software engineering companies. *ITM Web of Conferences, 6*.

Sutherland, J. (2014). *Scrum, the art of doing twice the work in half the time.* New York: Crown.

Swink, M. J. (2012). Six Sigma adoption: Operating performance impacts and contextual drivers of success. *Journal of Operations Management, 30*(3), 437-453.

Tabachnick, B. &. (2007). *Using Multivariate Statistics* (5th ed.). San Francisco, CA: Pearson.

Takeuchi, H. a. (1986, January). The New New Product Development Game. *Harvard Business Review*.

Tamura, R. B.-S. (1992). The use of repeated measures analyses in developmental toxicology studies. *Neurotoxicology and Teratology, 14*, 205-210.

Tanner, M. a. (2017). The Use of Kanban to Alleviate Collaboration and Communication Challenges of Global Software Development. *Issues in Informing Science and Information Technology, 14*, 177-197.

Taylor, A. (2011). *Using the GLM Procedure in SPSS.* Retrieved May 15, 2017, from Maqurarie University Psychology: https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=0ahUKEwid5-vYg6rUAhXMJiYKHcFICHAQFggsMAA&url=http%3A%2F%2Fwww.psy.mq.edu.au%2Fpsystat%2Fdocuments%2FGLMSPSS.pdf&usg=AFQjCNEi0zjqJ016P7gVl0nAVili46J1pw

Taylor, F. (1911). *The principles of scientific management.* New York: Harper & Brothers.

Taylor, W. (2000). *A pattern test for distinguishing between autoregressive and mean shift data.* Retrieved from Variation.com: http://www.variation.com/cpa/tech/pattern.html

Taylor, W. (2000, Sept 7). *Change Point Analysis. A Powerful New Tool For Detecting Changes.* Retrieved Oct 19, 2017, from Variation.com: http://www.variation.com/cpa/tech/changepoint.html

Thun, J.-H. D. (2010). Empowering Kanban through TPS principles- an empirical analysis of the Toyota Production System. *International Journal of Production Research, 48*(23), 7089-7106.

Tonelli, A. O. (2013, March 1). Improving the Management of Cost and Scope in Software Projects using Agile Practices. *International Journal of Computer Science and Information Technology*, 47-63.

Trading Economics. (2017, Nov 13). *United Kingdom Inflation Rate 1989-2018*. Retrieved from Trading Economics: https://tradingeconomics.com/united-kingdom/inflation-cpi

Trojanowska, J. D. (2017). Application of the Theory of Constraints for Project management. *Management and Production Engineering Review, 8*(3), 87-95.

Turetken, O. s. (2016). Assessing the adoption level of scaled agile development: a maturity model for Scaled Agile Framework. *Software Evolution and Process*.

UC Denver. (n.d.). *Repeated Measures ANOVA in SPSS.* Retrieved March 11, 2017, from UC Denver College of Nursing: http://www.ucdenver.edu/academics/colleges/nursing/Documents/PDF/RepeatedMeasuresANOVA.pdf

Varhol, P. (n.d.). *To Agility and Beyond: The history-and legacy-of agile development.* Retrieved August 15, 2017, from TechBeacon: https://techbeacon.com/agility-beyond-history%E2%80%94-legacy%E2%80%94-agile-development

Verma, R. (1997, April). Management Science, theory of constraints/optimized production technology and local optimization. *Omega, 25*(2), 189-200.

Version One. (2016). *10th Annual State of Agile Report.* Version One. Retrieved Dec 14, 2016, from https://versionone.com/pdf/VersionOne-10th-Annual-State-of-Agile-Report.pdf

Walenta, T. (2015, April 12). *PMI's Project Management Body of Knowledge process flow is iterative and incremental - and supports both agile and waterfall.* Retrieved Sept 14, 2017, from Linkedin Pulse:

https://www.linkedin.com/pulse/pmis-pmbok-process-flow-iterative-incremental-walenta-pmi-fellow/

Wallin-Miller, G. L. (2016, Aug). Anabolic-androgenic steroids decrease dendritic spine density in the nucleus accumbens of male rats. *Neuroscience, 330*, 72-78.

Wang, T. (2007). Comparison of Methods for Valuating Technology Innovation and Adoption Projects. *Dissertation, George Washington University*. Retrieved from file:///C:/Users/Andrew/Downloads/out%20(19).pdf

Wayhan, V. B. (2007). TQM and Financial Performance: what has empirical research discovered? *Total Quality Management, 8*(4), 403-412.

Weaver, P. (2007, June). A Brief History of Project Management. *Project, 19*(11), 9-12.

Wei, Z. J. (2014). Organizational ambidexterity, market orientation, and firm performance. *Journal of Engineering and Technology Management, 33*, 134-153.

Wilkinson, L. (1999). Task Force on Statistical Inference. *American Psychologist, 54*(8), 594-604.

Woan-Yuh, L. C.-I. (2008). An integrated framework for ISO9000 motivation depth of ISO implementation and firm performance. *Journal of Manufacturing Technology Management, 19*(2), 194-216.

Woodward, E. (2013, Aug 11). *Controversy around SAFe, DAD, and Enterprise Scrum*. Retrieved Sept 23, 2017, from IBM Developer Community Blogs: https://www.ibm.com/developerworks/community/blogs/c914709e-8097-4537-92ef-8982fc416138/entry/august_11_2013_8_56_am?lang=en

Wu, L.-Y. W.-J. (2007). Transforming resources to improve performance of technology based firms: a Taiwanese Empirical Study. *Journal of Engineering and Technology Management, 24*, 251-261.

York, K. M. (2004). Causation or covariation: an empirical re-examination of the link between TQM and financial performance. *Journal of Operations Management, 22*(3), 291-311.

Yunis, M. J. (2013). TQM strategy, and performance: a firm level analysis. *International journal of quality and reliability management, 30*(6), 690-714.

Zhang, G. P. (2012). Does Quality Still Pay? A Reexamination of the relationship between effective quality management and firm performance. *Production and Operations Management*, 120-136.

Zhang, M. (2005, July 18). Information Systems, strategic flexibility, and firm performance: an empirical investigation. *Journal of Engineering and Technology Management, 22*, 163-184.

Appendix I. Data summary.

Independent VariablesOrganization	Primary methodology	Year transitioned	Mkt cap size	Economic Environment	Age of Firm	Geography	Industry
BMC	SAFe	2005	Mid	0	3	0	0
CSG International	SAFe	2007	Sm	1	3	0	1
SEI Global Wealth Services	SAFe	2013	Mid	0	3	0	1
Salesforce	Scrum	2007	Lg	1	1	0	0
DST Systems	Scrum	2009	Mid	1	3	0	1
ASOS	Scrum	2011	Sm	0	2	1	2
Yahoo	Scrum	2005	Lg	0	1	0	1
Italtel	LeSS	2012	Sm	0	3	2	3
Paypal	Scrum	2013	Lg	0	2	0	1
IDX Systems	Scrum	1996	Sm	0	3	0	1
Google	Scrum	2005	Mega	0	1	0	0
Amdocs	SAFe	2014	Mid	0	3	3	3
TomTom	SAFe	2014	Sm	0	3	2	4
Nemetscek	Kanban	2010	Mid	0	3	2	0
Ultimate SW	Scrum	2005	Mid	0	2	0	0
US	Scrumban	2008	Mid	1	2	0	0
Ult	Kanban	2013	Mid	0	3	0	0
Bazaarvoice	Kanban	2014	Sm	0	1	0	2
Gogo air	SAFe	2012	Sm	0	2	0	3
Bottomline	Scrum	2011	Sm	0	3	0	1
Microsoft	Scrum	2011	Mega	0	3	0	0
Barclays	DAD	2014	Lg	0	3	1	5
Borland	Scrum	2008	Sm	1	3	0	0
John Deere	SAFe	2011	Lg	0	3	0	6
usg	Scrum	2015	Mid	0	3	0	6
Travis Perkins	SAFe	2013	Mid	0	3	1	6
systematic	Scrum	2006	Sm	0	3	3	0
HR Block	Scrum	2011	Mid	0	3	0	1
Trimble	scrum	2008	Mid	1	3	0	4
Ing	scrum	2011	Lg	0	2	2	5
sk hynix	SAFe	2014	Mega	0	3	4	4

Legend

Economic Environment 0 Bull

1 Bear

Age of Firm	1	less than 10 years
	2	10-20 years
	3	over 20 years
Geography	0	US
	1	UK
	2	EU
	3	multinational
	4	Korea
Industry	0	Software
	1	Business Services
	2	Retail
	3	Telecom
	4	Consumer Electronics
	5	Banking and Finance
	6	Industrial, Construction, Heavy Equipment

OER

Organization	-4	-3	-2	-1	0	1	2	3
BMC	1.005964	1.2205	0.98416	1.06982	0.9836	0.9364	0.86899	0.79376
CSG International	1.016858	0.72546	0.79743	0.77429	0.8	0.8108	0.85072	0.86468
SEI Global Wealth Services	0.758491	0.75852	0.78051	0.78684	0.7794	0.7213	0.73152	0.73305
Salesforce	1.002687	0.78125	0.77273	0.71521	0.7666	0.7423	0.73605	0.71415
DST Systems	0.866402	0.86353	0.851	0.84858	0.8765	0.8518	0.89112	0.93905
ASOS	0.927952	0.93023	0.92593	0.91515	0.9103	0.9147	0.91498	0.92848
Yahoo	0.597297	0.82566	0.73662	0.59754	0.4317	0.3905	0.43704	0.4929
Italtel		1.41224	1.03994	1.31797	1.0477	1.0745	1.03566	1.01235
Paypal		0.88883	0.87642	0.84458	0.8378	0.8375	0.83696	0.85185
IDX Systems		0.95726	0.93706	0.89143	0.8932	0.9562	0.85358	
Google	0.872093	0.57727	0.76655	0.79931	0.5847	0.6653	0.69362	0.69577
Amdocs	0.871768	0.86227	0.9012	0.86387	0.8562	0.8605	0.85836	0.87009
TomTom	0.493919	0.48915	0.50275	0.52507	0.541	0.5505	0.51539	0.57345
Nemetscek	0.859813	0.85616	0.88	0.84892	0.8235	0.8274	0.83799	0.83243
Ultimate SW	0.695	0.764	0.700	0.639	0.568	0.5439	0.52318	0.56742
US	0.695	0.764	0.700	0.639	0.5674	0.551	0.52863	0.50558
Ult	0.695	0.764	0.700	0.639	0.4829	0.4891	0.54369	0.5621
Bazaarvoice	0.815789	0.922	0.861	1.003	0.994	0.7958	0.73367	0.66
Gogo air		4.014	1.81915	1.20625	1.1159	1.1341	1.125	1.08782
Bottomline	0.644068	0.60305	0.64493	0.52866	0.4921	0.5134	0.55906	0.57333
Microsoft	0.637651	0.62774	0.65517	0.6129	0.3857	0.3784	0.39744	0.37931
Barclays	0.645161	0.73684	0.744	0.71212	0.6818	0.6227	0.67925	
Borland	0.796117	0.91304	0.94426	0.95755	0.907			
John Deere	0.888921	0.89014	0.94206	0.88367	0.8681	0.869	0.85493	0.86697
usg	1.07079	0.97736	0.90903	0.94869	0.8781	0.8694		
travis perkins	0.938567	0.93782	0.93681	0.93291	0.9324	0.9312	0.93066	0.93421
systematic	0.8901	0.97402	0.97707	0.961	0.8993	0.8994	0.9701	0.90046
HR Block	0.628302	0.63338	0.63556	0.60385	0.6031	0.588	0.52668	0.51984
Trimble	0.357784	0.34194	0.34681	0.35516	0.3484	0.4121	0.39985	0.40937
Ing	0.612613	0.56977	0.60241	0.60976	0.1846	0.1964	0.23077	0.21569
sk hynix	0.975841	0.75425	0.96873	1.02234	0.7614	0.7017	0.71614	0.81369

ROA

Organization	-4	-3	-2	-1	0	1	2	3
BMC	0.0138	-0.069	0.0158	-0.009	0.023	0.032	0.066	0.0936
CSG International	-0.0363	0.0664	0.0834	0.0915	0.144	0.111	0.0771	0.0255
SEI Global	0.1776	0.1694	0.1583	0.1588	0.2	0.207	0.2088	0.2035
Salesforce	-0.0357	0.026	0.0326	0.0007	0.019	0.042	0.0578	0.0579
DST Systems	0.1416	0.1042	0.1152	0.1236	0.122	0.143	0.0799	0.153
ASOS	0.14	0.1875	0.1629	0.1765	0.244	0.213	0.2216	0.2563
Yahoo	0.0309	-0.039	0.0151	0.04	0.091	0.175	0.0652	0.0523
Italtel		-0.255	-0.018	-0.259	-0.024	-0.06	-0.078	-0.089
Paypal			0.0413	0.0481	0.05	0.019	0.0425	
IDX Systems		0.0979	0.0689	0.0748	0.093	0.03	0.1056	
Google	0.3462	0.1206	0.1204	0.1426	0.167	0.166	0.1331	0.161
Amdocs	0.0753	0.0714	0.0746	0.0842	0.084	0.081	0.0838	0.0791
TomTom	0.0427	0.0417	0.0349	0.0062	0.022	0.018	0.0074	0.0084
Nemetscek	0.0686	0.0806	0.0659	0.0753	0.121	0.13	0.1273	0.1404
Ultimate SW	-0.2353	-0.452	-0.25	-0.096	0.049	0.043	0.2444	-0.014
US	-0.2353	-0.452	-0.25	-0.096	-0.014	-0.01	0.008	0.0126
Ult	-0.2353	-0.452	-0.25	-0.096	0.042	0.038	0.0157	0.026
Bazaarvoice	-0.2431	-0.526	-0.153	-0.141	-0.162	-0.1	-0.076	-0.05
Gogo air		-0.404	-0.326	-0.119	-0.063	-0.06	-0.065	-0.045
Bottomline	-0.037	-0.025	-0.066	0.0149	0.096	0.003	-0.024	-0.027
Microsoft	0.2222	0.2361	0.1795	0.2093	0.248	0.223	0.1966	0.157
Barclays	0.003	0.0025	-1E-04	0.0014	0.003	0.003	0.0027	
Borland	0.0221	-0.064	-0.117	-0.108	-0.01			
John Deere	0.0472	0.053	0.0212	0.0432	0.058	0.055	0.0594	0.0517
usg	-0.1048	-0.034	0.0116	0.0094	0.209	0.132		
travis perkins	0.0452	0.0385	0.0411	0.051	0.06	0.068	0.0753	0.0739
systematic	0.1033	0.0316	0.0516	0.0485	0.135	0.11	0.0931	0.1084
HR Block	-0.0808	-0.055	0.0905	0.0915	0.078	0.057	0.0957	0.1012
Trimble	0.1024	0.1121	0.1048	0.076	0.087	0.05	0.0552	0.0558
Ing	0.0084	-0.001	-0.001	0.0023	0.005	0.004	0.0037	0.0074
sk hynix	-0.0204	0.1432	-0.003	-0.009	0.135	0.156	0.1457	0.0919

Revenue Ratio

Organization	-4	-3	-2	-1	0	1	2	3
BMC	1.031442	0.88038	0.90636	0.96924	1	1	1.07997	1.18319
CSG International	0.897901	0.83755	0.89957	0.91388	1	1.126	1.19442	1.31059
SEI Global	0.941762	0.79994	0.82559	0.88135	1	1.1242	1.18477	1.24453
Salesforce	0.102577	0.19317	0.35481	0.62333	1	1.5061	2.16611	2.62641
DST Systems	1.134001	1.00807	1.03814	1.03043	1	1.0499	1.07701	1.16173
ASOS	0.080717	0.19283	0.36323	0.73991	1	1.5202	2.21973	3.45022
Yahoo	0.310576	0.20062	0.26665	0.45467	1	1.4709	1.79771	1.94992
Italtel	1.209302	1.04651	1.09044	1.16537	1	1.0207	1.09561	1.18605
Paypal	0.41549	0.52148	0.6688	0.84168	1	1.193	1.37476	1.61171
IDX Systems	0	0.56893	0.69513	0.84728	1	1.2153	1.5549	
Google	0.014011	0.07152	0.23868	0.51955	1	1.7276	2.70349	3.55099
Amdocs	0.85535	0.89181	0.94979	0.97041	1	1.0649	1.08876	1.11118
TomTom	1.536864	1.57944	1.32191	1.09761	1	0.9865	1.04465	1.02492
Nemetscek	0.718121	0.97987	1.00671	0.90604	1	1.1007	1.1745	1.24161
Ultimate SW	0.670	0.625	0.682	0.818	1.000	1.2955	1.70455	
US	0.404	0.494	0.640	0.843	1	1.1011	1.27528	1.51124
Ult	0.478	0.554	0.656	0.810	1	1.2317	1.50732	1.90488
Bazaarvoice	0.22619	0.381	0.631	0.869	1	1.1369	1.18452	1.19643
Gogo air	0	0.158	0.40598	0.67347	1	1.4017	1.74359	2.14103
Bottomline	0.624339	0.69312	0.73016	0.83069	1	1.1852	1.34921	1.59259
Microsoft	0.073231	0.86385	0.82925	0.89336	1	1.054	1.11303	1.24148
Barclays	1.273489	1.27697	0.98897	1.10467	1	1.0276	0.97975	
Borland	1.796512	1.60465	1.76744	1.22674	1			
John Deere	0.75228	0.88832	0.72198	0.81232	1	1.1295	1.18065	1.12667
usg	0.774459	0.85616	0.95846	0.99691	1	1.0357		
travis perkins	0.569153	0.61228	0.92832	0.94114	1	1.0839	1.15423	1.20765
systematic	0.688889	0.64444	0.64444	0.77778	1	1.1111	1.04444	1.06667
HR Block	0.973166	1.10088	1.09986	1.02378	1	0.9827	0.98675	1.02717
Trimble	0.503386	0.58239	0.7073	0.91949	1	0.8473	0.97291	1.23702
Ing	1.704712	1.32808	1.27561	1.25583	1	0.8614	0.80519	0.79406
sk hynix	0.558136	0.85464	0.73385	0.7174	1	1.209	1.32707	1.21405

Rejected companies

Organization	Method	reason rejected
Seamless	Scrum	couldn't confirm when change implemented
Tradestation	SAFe	private, no financials available
Valpak	SAFe	private, no financials available
bwin.party	LeSS	private, no financials available
Tableau	Drive	private, no financials available
Valve	Scrum	Cannot confirm date of transformation
Guidewire	cannot confirm	cannot confirm
Vodaphone	Scrum	Unable to get adequate financial data
Spotify	Scrum, its own	private at the time of transition
Atlassian	multiple	cannot identify transition time
Foursquare	scrum	private, no financials available
Etsy	scrum	private, no financials available
QSR International	RUP, SAFe	transition only from Agile to Agile
Lockheed Martin	SAFe	only small portion of organization impacted
NASA	SAFe	adequate financial data not available, only portion of organization impacted
Elbit Systems	SAFe	private, no financials available
Capital One	SAFe	only small portion of organization impacted
Deutsche Bank	SAFe	only small portion of organization impacted
NextGear Capital	SAFe	private, no financials available
NICE	SAFe	private, no financials available
Dutch Tax Authority	SAFe	adequate financial data not available, only portion of organization impacted
US Air Force	SAFe	adequate financial data not available, only portion of organization impacted

USPS	SAFe	adequate financial data not available, only portion of organization impacted
Fitbit	SAFe	adequate financial data not available, only portion of organization impacted
US Immigration	SAFe	adequate financial data not available, only portion of organization impacted
Northwestern Mutual	SAFe	only small portion of organization impacted
Philips	SAFe	only small portion of organization impacted
HP	SAFe	only small portion of organization impacted
Swisscom	SAFe	only small portion of organization impacted
Cisco	SAFe	only small portion of organization impacted
pole emloi	SAFe	adequate financial data not available, only portion of organization impacted
LEGO	SAFe	only small portion of organization impacted
Accenture	SAFe	only small portion of organization impacted
RMIT University	SAFe	adequate financial data not available, only portion of organization impacted
Intel	SAFe	only small portion of organization impacted
BMW	LeSS	only small portion of organization impacted
JP Morgan Chase	LeSS	only small portion of organization impacted
Alcatel Lucent	LeSS	only small portion of organization impacted
Ericsson	LeSS	only small portion of organization impacted
AFGA Healthcare	LeSS	couldn't confirm when change implemented

Openlink	DAD	private, no financials available
Panera	DAD	only small portion of organization impacted
Primavera	Scrum	private, no financials available
IMVU	Scrum	private, no financials available
GE	Scrum	only small portion of organization impacted
BBC	Scrum	only small portion of organization impacted
Schneider Electric	Scrum	only small portion of organization impacted
FBI Sentinel	Scrum	only small portion of organization impacted
Dutch Railways	Scrum	only small portion of organization impacted