The Development and Testing of a Contextual Model for Healthcare Quality Improvement using Lean and the Model for Understanding Success in Quality (MUSIQ)

A thesis presented in partial fulfilment of the requirements for the degree of

Master of Quality Systems

at Massey University, Palmerston North, New Zealand

William James Wilson

2017

Acknowledgements

I am indebted to my research supervisors, Dr Nihal Jayamaha and Dr Greg Frater of Massey University, for their regular advice and consistent support throughout this research. Both Nihal and Greg have provided invaluable feedback over the course of the project, from the initial planning and literature review through to the structure of the final presentation. Nihal also provided guidance in my technical analysis (e.g. how conceptual frameworks are transformed into hypotheses and how they could be tested) by providing me the initial sketches, directing me to appropriate reference sources on quantitative research, software platforms (I was only familiar with R, which was not very user-friendly in partial least squares path modelling) and providing technical explanations to some results that I encountered in my statistical analysis.

I am also grateful to my Canterbury District Health Board colleagues Lynn Davies, Brian Dolan, Ross Denton, Heather Manson, Tina Wright, Roxanne Mckerras and Yvonne Williams for their support and advice particularly getting started and pre-testing the survey but also at various latter stages of the research. Finally I wish to thank my wife Helen for her patient and honest reviews of the many early drafts of this document, and her unfailing support and enthusiasm for the entire project.

29 June

Canterbury District Health Board

Christchurch

New Zealand

Abstract

This study developed a new theoretical model of quality improvement (QI) contextual factors, for QI activity undertaken at the healthcare microsystem level. The *Model for Understanding Success in Quality* (MUSIQ) (Kaplan, Provost, Froehle, & Margolis, 2012), was aligned with Lean improvement activity using the Toyota Way framework. The aim of the research was to improve the effectiveness of healthcare quality improvement initiatives by providing more understanding of the associations, relative importance and precise functioning of critical contextual factors. A new survey instrument, based on the literature, was developed to collect data and the hypothesised theoretical relationships were tested using the partial least squares path modelling (PLSPM) technique.

QI practitioners at a large New Zealand District Health Board were surveyed on a range of contextual factors hypothesised to influence improvement outcomes. All survey participants had recently completed a small-scale improvement project using Lean, or were participants in training programmes that introduced them to Lean thinking and methods. Some participants worked autonomously on improvements of their own selection; others were part of a wider training programme derived from the National Health Service's (UK) 'productive ward' programme. In the healthcare organisational context, the majority of these improvement initiatives were carried out at the microsystem level – initiated and delivered by the teams responsible for the work processes being modified.

Survey responses were first analysed via principal components analysis (to examine the dimensionality of the scales) and then PLSPM. The defined contextual factors for 'Teamwork', 'Respect for People', 'Lean Actions' and the influence of negatively motivating factors all reached significance. Defined contextual factors for 'Previous Experience' and the influence of positive motivating factors did not reach significance at 5% level. The final model showed a statistically significant, moderate predictive strength, with an overall adjusted R^2 of 0.58. This result was an encouraging validation of the microsystem-level layer of the MUSIQ model using Lean as the QI method (context). The relative influence of 'Teamwork', 'Respect for People', 'Motivation', and a mediating mechanism for making process changes (in this instance, Lean) were measured and found to be consistent with the MUSIQ model. Identifying more detailed causal mechanisms (the present model was intentionally parsimonious due to the time

frame allowed and the resources available for the research), refining the operational definitions, and developing and testing predictive models for the defined contextual factors are the proposed next steps in the research.

List of Acronyms

AVE	Average Variance Extracted
BEF	Business Excellence Framework
CBSEM	Covariance Based Structural Equation Modelling
CDHB	Canterbury District Health Board
CFIR	Consolidated Framework for Implementation Research
CI	Continuous Improvement
CSF	Critical Success Factor
DHB	District Health Board
HR	Human Resources
MOH	Ministry of Health (New Zealand)
MUSIQ	Model for Understanding Success in Quality
NHS	National Health Service (United Kingdom)
PCA	Principal Components Analysis
PCR	Principal Components Regression
PLSPM	Partial Least Squares Path Modelling
QI	Quality Improvement
SEM	Structural Equation Modelling
TPS	Toyota Production System
TW	Toyota Way

Contents

Tables and Figures

List of Tables

List of Figures

Chapter 1
INTRODUCTION

1.1 Background

The public health system in New Zealand faces ongoing challenges to the delivery of high quality and sustainable healthcare. In terms of demand, there is continually increasing pressure due to population and disease trends. The disease burden is increasing due to an aging population, alongside increasing rates of cancer, obesity and diabetes (Mays, 2013). There are also constraints on the supply of care — as costs increase within fixed budget limitations (Ministry of Health, 2014). Unsurprisingly, there is a desire from all parties responsible for providing healthcare — the Government (via the Treasury and the Ministry of Health), district health boards (DHBs); and medical, nursing, and allied health professionals — to optimise healthcare value from limited resources.

In response to this pressure, one strategy that has been prominent in healthcare internationally over the last two decades has been the introduction of quality improvement (QI) methodologies, including Lean thinking, Six Sigma, and process re-engineering (Andersen, Røvik, & Ingebrigsten, 2014; Rubenstein et al., 2014). As is the case for other industry sectors, the introduction of QI methodologies in healthcare has the twin goals of improvement of service delivery and outcomes alongside maximising the value from constrained budgets (Brandão de Souza & Pidd, 2011). Since Lean thinking is a philosophy and a management practice that prompts the organisation to continuously improve the processes and outcomes to minimise nonvalue adding activities and waste from every aspect of operation (Shah & Ward, 2007; Womack & Jones, 2003), it is of little surprise that Lean as a quality and process improvement approach has attracted more attention from healthcare administrators than alternative improvement methodologies.

New Zealand is one of several countries which have adopted Lean healthcare models from the UK National Health Service, including the Lean-derived *Productive Ward* programmes. These programmes include fundamental Lean methods to minimise waste and identify key value streams as they relate to patient care (White, 2015). The introduction of Lean and Lean-derived quality improvement methods has encountered some resistance from health professionals, usually relating to concerns over the applicability in the healthcare setting, or challenges to the amount and type of evidence that the methods actually work (Moraros, Lemstra, & Nwankwo, 2016; Walshe, 2007).

A large volume of case studies and anecdotal evidence is now available to support the notion that Lean methodologies are effective in healthcare. However, given that case studies are not intended to be statistically generalisable, and that contextual factors are not always precisely defined, it can be difficult to generalise these findings to the wider population of healthcare organisations, either statistically or analytically (Andersen et al., 2014; Kaplan et al., 2010; Shojania & Grimshaw, 2005). In addition, the literature lacks studies on the mechanisms for achieving effectiveness (success), which remains highly context-dependent and poorly explained and measured (Andersen et al., 2014; Moraros et al., 2016). Continuing empirical research into the detailed enabling mechanisms, and the contextual requirements for successfully implementing quality tools is called for (Andersen et al., 2014; Kaplan et al., 2010; Moraros et al., 2016; Portela, Pronovost, Woodcock, Carter & Dixon-Woods, 2015). This empirical study, based on the secondary healthcare sector in New Zealand, aims to make a theoretical contribution to academia and a practical contribution to healthcare QI practice by attempting to fill the above void.

Alongside these higher-level societal and health sector requirements for a more complete picture, healthcare professionals and QI practitioners working in healthcare will also benefit from continued empirical research into what works and why. Tools and techniques that have been fine tuned in specific healthcare environments (for this study, secondary care clinical microsystems), will improve efficiency, increase the yield of successful QI initiatives and increase health professionals' confidence in the underlying methodologies, such as Lean. From the perspective of practitioners (the researcher is a QI professional with experience across multiple industry sectors), the search for simple tools able to be used independently by operational staff remains a work in progress. As will be seen from the literature review that follows, high level critical success factors for QI in general (and Lean in particular), such as organisational leadership or culture, are well identified; unfortunately there is very limited opportunity for individuals to bring about change at that organisational level, in part due to structural reasons (e.g. bureaucracy) and in part due to lack of full understanding how success factors for quality improvement contribute towards successful outcomes in a particular context, within and outside healthcare (Gonzalez & Van Aken, 2016; Kaplan, Provost, Froehle, & Margolis, 2012; Taylor et al., 2013). However, individuals can lead and participate in effective QI activity. It is in this second aspect that this research aims to create an impact.

QI programmes and methods are now a cornerstone of public health strategy and practice in New Zealand (Ministry of Health, 2012). QI programmes are standard and expected activity within New Zealand DHBs — the regional autonomous public entities responsible for the provision of secondary and tertiary healthcare. It is therefore appropriate to take a closer look at practical quality improvement efforts within this New Zealand secondary healthcare setting, with a focus on understanding the role 'context' plays in improvement outcomes.

1.2 The New Zealand Secondary Healthcare Sector

The New Zealand public health system is predominantly funded from central government funding via taxation (referred to as Vote Health). The size of this funding is approximately 9% of GDP (Ministry of Health, 2017). Funding for treatment resulting from accidents is provided by the Accident Compensation Corporation scheme. Completing the picture, a smaller (approximately 20% of the total) private health market, funded predominantly by insurance, complements the public system (Ministry of Health, 2017). The majority of the Vote Health funding is allocated on a population and health needs assessment formula to 20 District Health Boards (DHBs), which plan, fund and provide health services to the populations in their geographic regions. DHBs also fund the Primary Health Organisations (PHOs), the primary care services that are usually the first contact point patients have with the health system. Almost all GP practices and medical centres are members of the 32 PHOs nationwide (Ministry of Health, 2017).

Figure 1.1 shows a simplified model of the relationship between the Ministry of Health, DHBs and PHOs. In addition to funding primary care services, DHBs also provide specialist secondary services via their provider arms – usually hospitals. Highly specialised services such as neurosurgery may only be available at tertiary hospitals, which also provide care for patients from smaller hospitals and other DHBs.

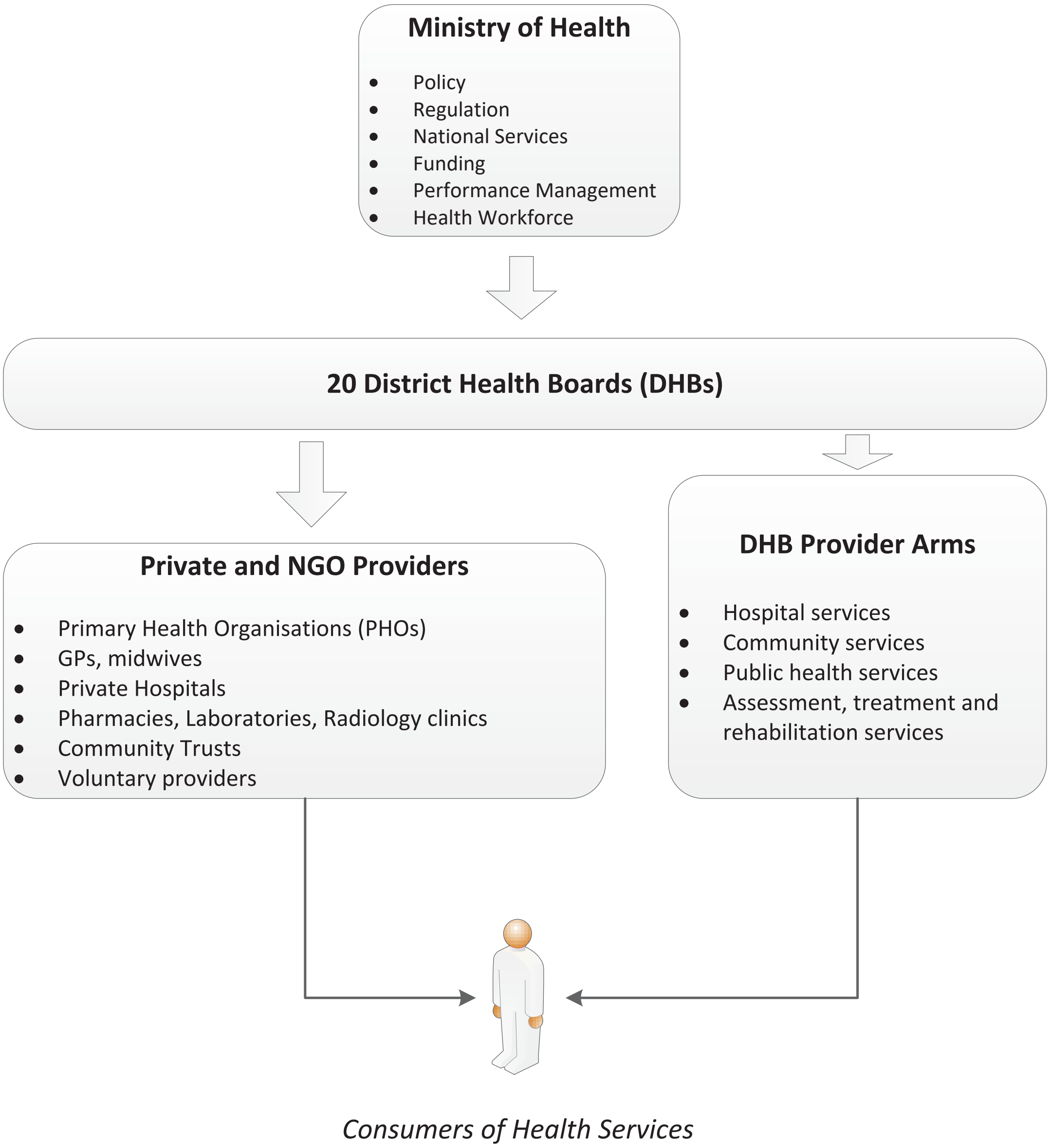


Figure 1.1: Simplified Structural View of New Zealand Health Service Providers *Adapted from "The structure of the NZ Health and Disability Sector"* (Ministry of Health, 2017).

The Canterbury District Health Board, the location for the present study, is the second largest DHB by population (Fig 1.2).

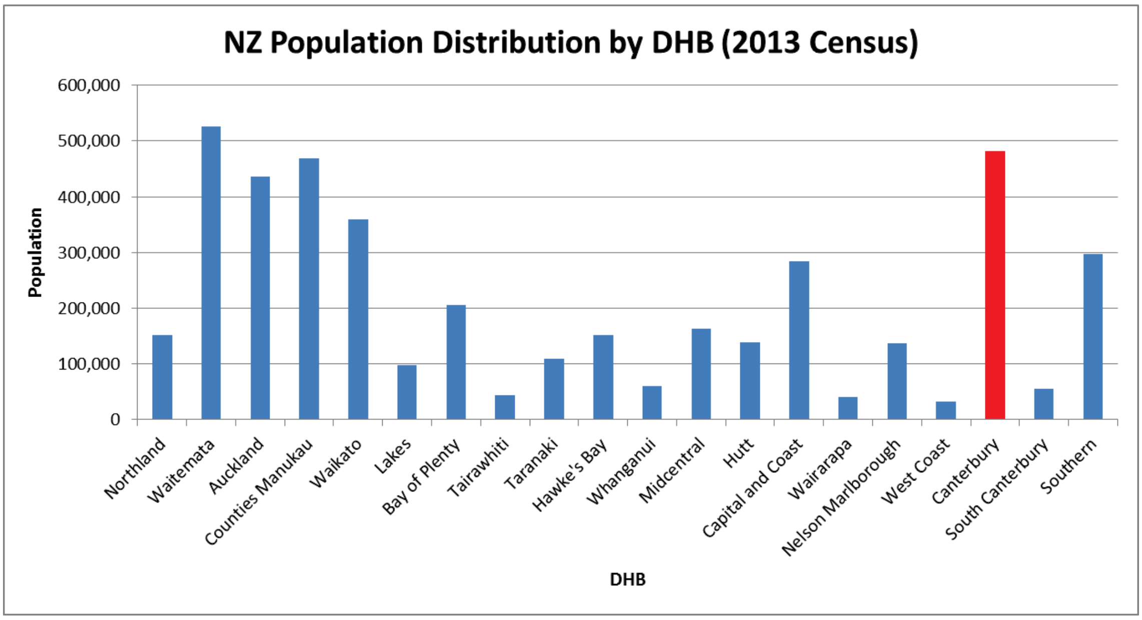


Figure 1.2: DHB Size by Population
Adapted from 2013 Census district health board tables (Statistics New Zealand, 2013).

As an indicator of how this DHB size translates into healthcare activity, the CDHB employs approximately 9500 people at six major hospitals and 30 rural and community facilities. A similar number of health professionals deliver services via the more than 1000 health service contracts administered by the CDHB. In the 2015/16 financial year, there were 21,039 elective (planned) surgical discharges from CDHB hospitals, 15,500 acute (emergency) surgeries, and 94,000 attendances at CDHB emergency departments. Approximately 580,000 outpatient appointments take place in Canterbury per year (Canterbury District Health Board, 2017).

The New Zealand public health system is comprehensive and generally well regarded; the World Health Organisation ranked the NZ health system 41 out 191 health systems reviewed in 2000 (World Health Organisation, 2000). However, the aging population, increasing disease burden and increasingly sophisticated and expensive interventions place pressure on the system. In New Zealand the most obvious manifestation of this pressure is access to services. Concerns over access and long waiting times for mental health services, major joint replacements and new cancer therapies have all been raised in the NZ media over the last 12 months. All the DHBs actively strive to deliver the necessary level of care to their populations safely and efficiently. Stretching the finite funding pool further by more effective use of QI methodologies, including Lean, is a priority for the sector.

1.3 Contingency Theory and Healthcare QI Context

The phrase "it depends", which is grounded in the general contingency theory of management, remains one of the most widely used phrases in explaining managerial phenomena (Borkowski, 2016; Luthans et al., 2015; Morgan, 2007). In essence, the general contingency of management posits that what works best or which action becomes more effective than another depends on the particular situation or the context in which a system or the controller of the system (i.e. the leader) operates.

A general contextual theory in understanding quality improvement in healthcare, proposed by Kaplan et al. (2012), — the Model for Understanding Success in Quality (MUSIQ), has created significant interest among healthcare quality improvement researchers and practitioners. However, a healthcare organisation may also operate (or attempt to operate) a Lean management system, which is also a context that needs to be considered when applying a generic contextual model such as MUSIQ (e.g. people in a Lean sociotechnical system rely on teamwork and respect). In addition, there are other contextual factors that can affect quality improvement effectiveness in a Lean sociotechnical system. Thus it is timely to undertake research that jointly examines these elements in order to advance understanding of Lean in healthcare.

1.4 The Research Problem, Aim and Objectives of the Study

1.4.1 The Research Problem and the Aim

Given the above background, the research problem is to develop a testable theoretical model that explains how contextual factors affect the effectiveness (success) of QI initiatives in a Lean sociotechnical system. Thus, the aim of the study is to improve the effectiveness of healthcare quality improvement initiatives by providing more understanding of the associations, relative importance and precise functioning of critical contextual factors.

1.4.2 The Research Objectives

The specific objectives of the research are as follows:

1. To identify and measure the prevalence of contextual factors that affect quality improvement initiatives in New Zealand public hospitals.

2. To explain the key empirical relationships between quality and process improvement interventions, outcomes and the contextual factors, from a theoretical and practical standpoint.

The general objective of the study is to propose guidelines for quality improvement in the secondary healthcare sector, based on the empirical study findings.

1.5 The Research Questions

The research questions (derived from the literature review) that underpin the study are:

RQ1: What are the specific contextual factors that influence the success of quality improvement initiatives within a New Zealand secondary public healthcare context?

RQ2: To what extent do specific contextual factors enable or inhibit the success of quality improvement initiatives within a New Zealand secondary public healthcare context?

1.6 Methods Overview

Having developed a theoretical model based on the extant literature (details in Chapter 3), the researcher operationalised the model, collected data, tested the data and then finally tested the model. Using literature, a survey questionnaire using a five-point Likert type scale was developed to operationalise the constructs of the theoretical model. The constructs were treated as latent variables that require multiple measurement items to reflect each construct. The hypothesised theoretical relationships between the constructs (latent variables) were tested using the latent variable path modelling technique “partial least squares”, using Smart PLS 3.0 software, based on the survey data collected from a sample of respondents (n = 105), belonging to the Canterbury DHB (CDHB). Alongside the Partial Least Squares Path Modelling (PLSPM) technique, Principal components analysis (PCA) was performed to test unidimensionality of measurement scales and method bias. These tasks were performed using the freeware statistical analysis package R studio.

1.7 Limitations of the Study

To keep the statistical modelling relatively parsimonious, the study restricted the number of contextual factors measured. In addition, the contextual factors were assumed to be independent of the causal factors (the Lean sociotechnical system). The correlations between the contextual factors and the causal factors suggested that this assumption may be questionable.

Approximation of Lean management systems (more precisely, the Lean sociotechnical system) in hospitals to the Toyota Way is the second limitation (the comparison between the two sociotechnical systems is given later).

The third limitation is that the study findings (the results of the data analysis) are based on a convenience sample; a nonprobability sample belonging to respondents attached to the CDHB.

The fourth and final limitation is that the sample size is small (n = 105), which results in a high probability of missing *small moderation effects* involving the contextual factors (i.e. low statistical power). It is argued that small moderation effects are of limited importance to the practitioner; hence although a small sample size affects the p values associated with the moderation hypotheses, this will not materially affect the interpretation of the results from a practical perspective. It is important to note that resource limitations (funding and time) contributed to the third and fourth limitations.

1.7 Structure of the Thesis

The rest of the thesis is organised as follows.

Chapter Two outlines the literature relevant to the research problem. The literature is used to define key terms and the background of Lean management in Healthcare is explored. The problem of context in relation to QI effectiveness and the development of contextual models is reviewed.

Chapter Three covers the derivation of the theoretical model (hence the hypotheses) from the extant literature. This details the process of aligning two existing theoretical models to provide a plausible foundation for Lean as a healthcare QI methodology. The model is presented as a causal-predictive model involving seven latent variables of varying degrees of abstractness. The purpose of this model is to explain and predict success in quality improvement, using the Lean sociotechnical elements and other contextual factors that operate within this sociotechnical system. The rationale and support from theory for motivation as a moderating variable concludes the chapter.

Chapter Four covers the methodology adopted in the study (the researcher used a positivistic paradigm). Research strategy, sampling and survey design are described. Then the selection of measurement items to operationalise the latent variables of the causal model is described in detail. Justification of the use of the data analysis technique partial least squares path modelling (PLSPM) to test the hypotheses, in favour of the more widely used covariance based path modelling, is also given in this chapter.

Chapter Five covers the results of the quantitative data analysis. Following preliminary data checks, the results of PCA and PLSPM are described in detail. This section covers the estimated parameters of the modelling and the specific reliability and validity tests conducted on the data. Acceptance criteria are defined and items requiring further comment beyond the nominal criteria are examined.

Chapter Six discusses the results in the light of the literature and practical implications of the findings. Each of the hypotheses is considered and the findings related to previous research noted in the literature. Direct effects, indirect effects and moderating effects are all discussed.

Finally, Chapter Seven concludes the study by summarising the key outcomes of the study and explaining how these outcomes achieved the aim and objectives of the study. The limitations of the study are then revisited and recommendations for future research are given.

Chapter Two

LITERATURE REVIEW

2.1 Introduction

This chapter covers the key literature pertaining to the research problem. The next section (section 2.2) begins with the definitions of key terms associated with the research problem. This is followed by a review of the literature on Lean management with a particular emphasis on healthcare (section 2.3). The next section (section 2.4), which is the central section of the literature review, examines contextual models in healthcare quality improvement (QI). Section 2.5 highlights the knowledge gap and the research questions posed to address this knowledge gap. Finally, section 2.6 summarises the key points/identified from the literature.

2.2 Definition of Key Terms

The terms 'contextual factors', 'quality improvement' and 'effectiveness/success' are central to the supporting theories and implementation of this research, so they are first explained using definitions from the literature.

2.2.1 Defining Contextual Factors

In social research, context factors or contextual factors are independent factors in the environment, whose relative presence or strength can increase or decrease (moderate) the intended causal effect or the intervention (Shalley & Gilson, 2004). This provides a starting definition, positioning contextual factors outside of, but influencing, a cause-effect relationship. In relation to QI, this cause-effect relationship is the improvement intervention/activity and the associated outcome. Øvretveit (2010) defines 'context factors' as *"all factors that are not part of a quality improvement intervention itself"* (Øvretveit 2010, p.18). Øvretveit further notes the difficulty of defining this clear boundary between context and intervention in some cases. It is not always easy to distinguish whether activity is part of the intervention or part of the circumstances or 'conditions for improvement', such as the availability of information technology (Øvretveit 2010). This problem highlights the need for precise and standardised descriptions of context for research to be more widely applicable.

Kaplan et al. (2012) define contextual factors in relation to QI as "*characteristics of the organisational setting, the environment, the individual, and their role in the* *organisation or quality improvement (QI) project. Contextual factors are distinct from the technical QI process"* (Kaplan et al., 2012, p.13). This definition is congruent with the conditions for improvement idea as distinguished from the intervention activity itself (the cause and effect relationship). A neutral terminology for context is preferable to other possible labels for contextual elements, such as enabler, inhibitor or facilitator as used by Andersen et al. (2014). The label "contextual factor" better emphasises the distinction from the intervention activity, and the same factor may be any of enabler, facilitator or inhibitor in different situations. Hovlid & Bukve (2014) refer to both Kaplan et al. and Øvretveit's definitions of context, providing a recent example where this definition of 'contextual factors' is in accepted use by researchers in this subject area.

A final consideration for contextual factors is whether or not they are an intrinsic component of the 'conditions for improvement' outlined above (e.g. organisational leadership), or a more temporary phenomenon such as an individual project team's communication norms. Worchel (1986) defines a contextual factor as a *transitionary factor* that interacts with the cause to change the outcome. This temporal distinction appears to be underdeveloped in the contextual models of both Kaplan et al. (2012), and Andersen et al. (2014). For this study, all the variables of interest fall under the broader definition of contextual factors as supplied by Øvretveit (2010) and Kaplan et al., (2012); but they are then further defined within the language of statistical modelling as exogenous or endogenous variables, and then specifically modelled as cause, mediator, moderator or outcome variables. The moderator variables equate to Worchel's (1986) and Shalley and Gilson's (2004) definitions of factors of influence outside the cause and effect relationship. This more precise labelling of a given contextual factor's presumed functional role in any examined cause-effect relationship is a suggested refinement to the application of the healthcare QI contextual factor models.

2.2.2 Defining Quality Improvement (QI)

QI activities are variously referred to in the literature as *projects*, *initiatives*, *interventions or collaboratives* (Damschroder et al., 2009; Portela et al., 2015). Portela et al. (2015) provide a thorough overview of improvement interventions, which they define simply as *"purposeful efforts to secure positive change"* (Portela et al., 2015, p.325). This is acknowledged as a very general description, but a wide-reaching definition is required to encompass the vast range of improvement activity in healthcare. QI projects are further distinguished by Portela et al. (2015) as being *"primarily undertaken to achieve an improvement, as opposed to research aimed at generating new knowledge"* (Portela et al., 2015, p.326). Thus, the *effectiveness/success* of the outcome is important to a QI intervention. For this study, intervention types of varying size, duration, complexity and intended objective were considered for inclusion, provided the activity met the criterion of *"purposeful efforts to secure positive change"* (additional selection criteria are described in the methodology section).

2.2.3 Defining Effectiveness/Success

Since this research undertakes empirical testing of a QI model — the Model for Understanding Success in Quality (MUSIQ, Kaplan et al., 2012), it is necessary to define success in this specific context. Kaplan et al. (2012) defined success as "*the implementation of system and process changes with associated outcome improvements*" (Kaplan et al., 2012, p.15). 'Success' is treated as being equivalent to effectiveness; that is, an improvement can be considered successful if it was effective in delivering the outcome improvements. QI 'Effectiveness/Success' is the endogenous (dependent) variable in this study. The measurement of the effectiveness/success variable is examined in the methodology section.

QI interventions may also have some training or skills development objectives, where staff can practice and then apply new learning to other interventions (Portela et al., 2015). Whilst this additional measure of QI success is acknowledged, it was not examined in this research.

2.3 Lean Management

To orient readers unfamiliar with Lean, a brief summary of the major milestones in the evolution of Lean and its subsequent spread from manufacturing to service industries, in particular healthcare, is provided here. The 1990 book by Womack Jones, and Roos entitled "The machine that changed the world" first introduced the world outside Japan to the term 'Lean' (Womack & Jones, 2003). However the roots of Lean are much deeper, stretching right back to the production lines developed by Henry Ford at the beginning of the twentieth century. Taiichi Ohno, who first documented the Toyota Production System (TPS), acknowledged both Henry Ford and the North American supermarket system in his development of 'Just in Time' thinking (Shah & Ward, 2007). "Lean" became the default North American and European label for the systems and tools making up the TPS (Holweg, 2006); supported by a system of management principles codified by Toyota as the Toyota Way (TW) (Liker, 2004). Whereas the TPS makes up the tangible and more readily transferrable components of Lean, The TW elements are the intangible (and therefore much harder to replicate) resources of the Toyota sociotechnical system (Jayamaha et al., 2014).

The label ‘Lean production’ directly reflects the TPS as a system of ‘production’. The major elements of this system evolved at Toyota over many decades. Chief amongst these is the concept of ‘Just in Time’ production — producing only what is needed in small production runs. The oil shocks of the 1970s necessitated radical changes in the automotive industry. ‘Just in Time’ production prevents overproduction and allows high responsiveness to demand fluctuations and varying consumer choice, a necessity if the auto manufacturers were to survive (Shah & Ward, 2007). At Toyota, alongside ‘Just in time’ techniques, Toyota developed a series of principles and tools focused on maximising quality and minimising waste. Prominent TPS tools include ‘Kanban’, a self-managing, signalling system for levelling supply chains via pull production to ensure sufficient (but no more) stock is on hand. ‘Heijunka’ is the process of levelling and optimising multiple production flows simultaneously (Graban, 2009; Shah & Ward, 2007). The technical tools are supported by management principles aimed at supporting the workforce, sustaining change and orienting the entire organisation to maximising value by eliminating waste (Kaplan, Patterson, Ching & Blackmore, 2014; Liker, 2004). ‘Kaizen’ is the process of continuous, incremental improvement facilitated by the production workers doing the actual work. ‘Jidoka’ is the principle of building quality into the process activity itself, philosophically very different from quality by inspection after production. Jidoka is supported by standardised work (eliminating variation) and ‘Poka-Yoke’, a process for eliminating errors if they are discovered (Graban, 2009). Many of the core tools of the TPS and the TW principles are simple and conceptually elegant, which has contributed to their appeal and spread beyond their originally designed functions on the production line.

In 1996 Womack and Jones published their book “Lean Thinking: banish waste and create wealth in your corporation”. This book introduced the label “Lean Thinking”, which positioned the strategies and concepts of the TPS (underpinned by the TW) as a management system applicable at an enterprise level and industry sector beyond production and manufacturing (Womack & Jones, 2003). The five fundamental components of Lean Thinking are: *Value (define value to your customer), Value Stream (identify the steps in your process that create value), Flow (refine your process to link value adding steps and eliminate waste, Pull (the consumer triggers the production/value creation), and Perfection (seek perfection via continuous improvement.* A sixth component *lean consumption* (Womack & Jones, 2005), was added later to emphasise the fact that the customer value proposition is not something that is attributable only to good production line operation (e.g. pull production, just in time, Kanban, and line balancing), but also to good product design, because value is either created or destroyed in the consumption stage, when the customer actually uses the product or consumes the service (Gamage, Jayamaha, & Grigg, 2016; Kollberg, Dahlgaard, & Brehmer, 2006).

The abstraction of the original production line principles to this management level allowed many non-manufacturing industries to understand and adopt Lean Thinking. Any process delivering products or services to consumers could conceivably benefit from the application of Lean Thinking, potentially offering organisations adopting it a competitive advantage (Womack & Jones, 2003). Lean Thinking spread beyond automotive manufacturing to other production based sectors and then began diffusing into service based organisations. In the case of airlines, for example, the complex logistics and supply chain connections are immediately apparent; however Lean thinking continued to be taken up by other service-based sectors where the connections to physical production are not so obvious, including Healthcare. Completing this transition from ‘Lean Production’ to ‘Lean Management’ the latter term usually refers to the combination of enterprise-level application of techniques based on the TPS, incorporating ‘Lean Thinking’ principles and the integrated management principles of the TW (Liker, 2004).

Although the gap between automotive manufacturing and healthcare delivery may seem very large, the healthcare sector interest in Lean is understandable when the high-level principles are kept in mind. Hospitals and other healthcare providers have a strong interest in minimising waste and ensuring high quality, as well as managing complex ‘just in time’ delivery of care. Beyond the high-level principles, the specific tools and methods are also highly adaptable and often completely industry neutral. A3 reporting, visual management and value stream mapping, for example are now usually considered as generic quality management tools rather than elements of the TPS/TW.

Within the hospital environment, the delivery of care to a patient is necessarily just in time, in that it can't be produced before it is required. The right combination of resources has to be brought together at the right time, a direct echo of Ohno's original goal of the right product, at the right time, in the right quantity (Shah & Ward, 2007). A centralised diagnostic service such as Radiology must balance multiple requests across multiple machines; effective systems for levelling the workload and balancing across different patient groups and priorities are required. Pharmacy and linen supplies must flow continuously. Patient safety requires clear and embedded protocols that are not subject to errors in individual interpretation. Level loading (Heijunka), Kanban and Standard Work are highly applicable concepts to this environment (Graban, 2009).

Successful Lean implementation is necessarily much more than the adoption of simple tools; it is a fundamental orientation of the organisation. Sociotechnical systems theory focuses on the links between human behaviour and technological systems. Joosten, Bongers and Janssen (2009), highlight the dominance to date of Lean interventions focussed at the operational level (e.g. the introduction of tools or single process optimisations) over larger sociotechnical interventions (fundamental transformation of organisations). One widely reported early adopter of Lean in Healthcare was the Virginia Mason Medical Centre in Seattle, which created its "Virginia Mason Production System", based on the TPS, in 2004. This was in response to increased patient safety concerns and severe financial pressure (Kaplan, Patterson, Ching & Blackmore, 2014). Reviewing the progress a decade later, Kaplan et al. are clear that the isolated introduction of tools from the Lean toolkit is insufficient, and a clear leadership strategy to effect genuine change to the sociotechnical system is required (Kaplan et al., 2014). The varying level of success to date for healthcare organisations attempting to make these adaptations is explored in the next section.

2.4 Contextual Models in Healthcare QI

The use of QI methodologies has been underway in healthcare for at least two decades (D'Andreamatteo, Ianni, Lega, & Sargiacomo, 2015; NIST, 2017; Shojania & Grimshaw, 2005). Of these QI methodologies, Lean in particular has been popular and many case studies have been compiled and systematically reviewed (Andersen et al., 2014; Moraros et al., 2016; Waring & Bishop, 2010). In many industry sectors, Lean is increasingly integrated with Six Sigma to combine the best of both methodologies (Antony, 2011), but this integration is still comparatively rare in healthcare. As explained in the previous section, Lean management primarily aims at maximising value by attempting to eliminate waste from every facet of the operations; Six Sigma on the other hand attempts to create value by reducing variation to unprecedented levels (Antony, 2011).

An issue raised by Andersen et al. (2014) concerning many of the case studies on Lean activity in healthcare is that the body of evidence is often problematic in terms of the rigour expected of scientific enquiry. There are many single case studies with poor generalisability, anecdotal and subjective reports, and vague definitions and measures (Andersen et al., 2014). Furthermore, even in well executed and reported case studies which can be generalised to theory, it is obvious that successful outcomes for quality initiatives are not guaranteed. As healthcare practitioners have been adopting these methodologies on a larger scale, multiple barriers to implementation have been identified (Brandão de Souza & Pidd, 2011).

Implementation challenges are by no means limited to healthcare; a significant theme within the general discourse on Lean is the identification of critical success factors (CSFs) for implementation. The ongoing effort to identify and manage these CSFs indicates that implementing Lean is not a straightforward exercise in any industry sector. Business excellence frameworks (BEFs) such as the Malcolm Baldridge Award criteria or the European Foundation for Quality Management model provide a common foundation for establishing the basic operational conditions necessary for any organisation to conduct quality improvement effectively (Sampaio, Saraiva, & Monteiro, 2012). Amongst BEFs, the Shingo model offers a framework where the required principles for successful implementation of Lean are explicitly identified, with a strong emphasis on having the cultural enablers in place before proceeding to specific continuous improvement tools (Machado Guimarães & Crespo de Carvalho, 2014).

A substantial body of case study research covering Lean implementations provides a complementary empirical base to the more theoretical perspective of the business excellence frameworks, identifying critical success factors to Lean and Lean Six Sigma implementations across manufacturing and service sectors (including healthcare). Laureani and Antony (2012); Sisson and Elshennawy (2015); and Sreedharan and Raju (2016), provide succinct, contemporary overviews of CSFs. These factors span different levels of intervention, from organisational fundamentals such as leadership and management commitment, to specific project deployment strategies such as training and tool selection (Laureani & Antony 2012). Although there are differences of emphasis, a consensus is clearly discernible, and the criticality of *leadership, culture, engagement and processes*, as summarised by Sisson & Elshennawy (2015) is unlikely to be challenged by academic researchers or practitioners.

Returning to healthcare, Maher, Gustafson and Evans (2008) place their emphasis on sustainability, focusing on the conditions required for sustaining change once it has been implemented. In effect their sustainability guide is a practical tool for monitoring the degree of staff engagement, leadership behaviour and the presence of monitoring processes. This is certainly useful, but it is really an operational response to the critical success factors already noted, rather than an extension of the underlying theory.

Brandão de Souza and Pidd (2011) compare implementation barriers to Lean across manufacturing and healthcare (within the UK National Health Service). They identify three unique differences in healthcare culture: perceptions of Lean (compared to perceptions from within manufacturing), the personal and professional skills of healthcare workers, and hierarchy and line management differences (Brandão de Souza & Pidd, 2011). Nevertheless, the barriers identified as common across the sectors outnumber the healthcare specific barriers, and include the business excellence fundamentals such as adequate data collection and coherent organisational strategy (Brandão de Souza & Pidd, 2011). It is reasonable to conclude that established models of business excellence would be beneficial for Lean implementations in healthcare. Necessary, perhaps, but not sufficient, as the literature suggests there are additional factors to consider.

The specific healthcare implementation barriers to Lean break down into two main subgroups: concerns over the applicability of Lean within the healthcare environment (Brandão de Souza & Pidd, 2011; Walshe, 2007; Young & McClean, 2008), and the problem of adequately distinguishing the context (e.g. critical success factors for quality improvement) from the intervention itself (Andersen et al., 2014; Hovlid & Bukve, 2014; Øvretveit, 2010). These two distinct threads in the literature are now examined in more depth.

2.4.1 Applicability of Lean in a Healthcare Setting

The issue of the applicability of Lean to healthcare itself subdivides into two strands: 1) the healthcare-specific cultural differences, and 2) the evidence that Lean actually works. The healthcare-specific cultural barriers are multiple. Healthcare delivery takes place within traditional organisational hierarchies and Lean (and other QI methodologies) can disrupt these hierarchies (Brandão de Souza & Pidd, 2011; Waring & Bishop, 2010) Without full understanding of the evolution and strength of the relationships that exist in multi-disciplinary teams, such as within clinical microsystems, a non-context sensitive QI intervention is likely to meet resistance (Waring & Bishop, 2010). As the name suggests, the microsystem is the smallest granular unit of a system (in this research, the public health secondary care 'system'). Nelson et al. (2008, p. 368) describe clinical microsystems as the "basic building blocks of health systems", consisting of patients, small teams of health practitioners, inputs and outputs, and the specific systems, processes and feedback loops involved in providing health care to patients.

Regarding healthcare organisations as complex systems opens up one avenue for objectively assessing Lean and potentially bridging the perceived gaps across different industries or sectors. In their conceptual paper, Saurin, Rooke, and Koskela (2013) have addressed exactly this; they conclude that Lean is highly compatible with complex systems theory (CST). However they also highlight five important areas where Lean can learn from CST. Two areas standout in relation to this study: a frequent lack of emphasis on soft skills and a lack of realism in training Lean concepts (Saurin et al., 2013). Both of these elements — soft skills and realistic scenario-based training, are important for helping people function well in environments of high stress, unpredictability and high consequence — such as healthcare (Saurin et al., 2013). Researchers of Lean Management are already alert to the key linkages between successful Lean implementations and what are described as 'soft' Lean Management practices. Bortolotti, Boscari, and Danese, (2015), in their comparison of manufacturing plants internationally, found that successful plants consistently demonstrate higher levels of soft practices concerning people and communication (such as small group problem solving), than less successful plants with comparable levels of 'hard practices'— the specific techniques and methods of Lean (Bortolotti et al., 2015).

Young and McClean (2008) provide a different cultural perspective on Lean with their discussion of *value* creation in relation to healthcare. 'Whose value is being created' is the critical question that they posed. They point out that unless clinical, operational and experiential dimensions of value are precisely defined and agreed upon by everyone involved in an improvement initiative, motives of Lean may be questioned, and assessment of success (i.e. achieving more of which desired value) will be open to dispute (Young & McClean, 2008).

Concerns over the evidence for Lean's effectiveness in healthcare relate to the over-emphasis on case studies in the literature; there is limited evidence from controlled, repeatable experimental trials — which are the established norm in the assessment of clinical effectiveness (Walshe, 2007). However, several commentators (e.g. Andersen et al., 2014; Kaplan et al., 2010; Walshe, 2007), have highlighted that quality improvement in healthcare is a "complex social intervention" (Walshe, 2007, p. 57), and therefore not well suited to experimental research design methods. Walshe does not advocate for the abandonment of experimental method altogether, but he argues that experimental designs must be complimented by a theoretical approach, as is the norm in other social science research such as education and justice (Walshe, 2007).

2.4.2 Existing Contextual Models

The underlying justification for a theoretical approach is that the interaction of factors for any given quality improvement initiative is so dynamic and complex that the results from each situation are not empirically generalisable. However a sufficient understanding of the context may be achieved so that results can be generalised to theory. The aim is not just to find out whether a method or intervention works, but *when, how* and *why* an intervention works (Walshe, 2007). Retrospective studies based on empirical testing of cause-effect theoretical models on the effectiveness of QI interventions in healthcare, particularly with the inclusion of contextual factors, provide a useful alternative to experimental designs (Andersen et al., 2014; Kaplan et al., 2010). Done well, this strategy will address the concerns over the quality of evidence showing Lean works in Healthcare settings *and* the lack of precision distinguishing context from intervention activity (aiding replication and wider take up of successful methods). The literature review shows that development of context-sensitive theory is well underway. The remaining decision for a researcher is to select or modify an existing model, or to start afresh.

Kaplan et al. (2012) began the development of their contextual model with a comprehensive literature review, and the results of this were then shared amongst an international panel of ten QI experts. This group then went through a formal collaboration process to analyse the results, and identify and define important contextual factors. The outcome of their work is called the *Model for Understanding Success in Quality* (MUSIQ).The intent of the developers is to use the model as a basis for explaining contextual factors and the interactions in detail. The emphasis is entirely on context and includes a system of hierarchical layering — contextual factors are considered at external, macro and micro levels. An example of this layering is leadership being identified as important at each of the organisational, QI team, and microsystem levels. The microsystem layer of the MUSIQ model is shown below as Figure 2.1.

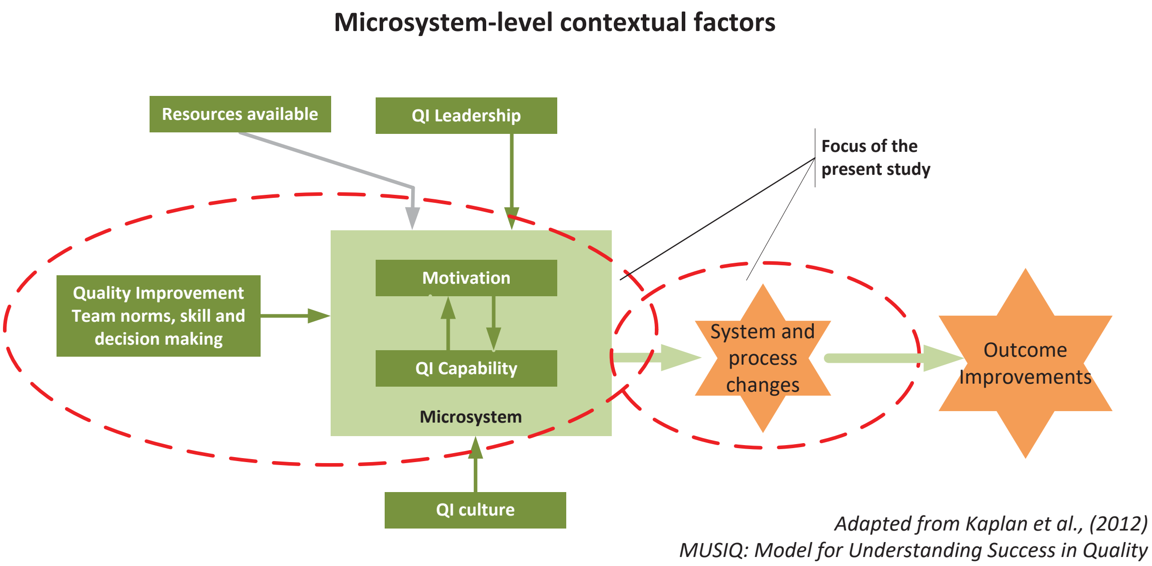


Figure 2.1: MUSIQ Model — Microsystem Level

Although still relatively new, various aspects of MUSIQ are actively being tested by researchers (Barson, Doolan-Noble, Gray, & Gauld, 2017; Griffin et al., 2017; Hovlid & Bukve, 2014; Kaplan et al., 2013). These studies all focus on identifying contextual factors as well as trying to understand how each factor affects quality improvement. The interest from researchers is encouraging, and suggests that MUSIQ may be a useful framework for these investigations, with potentially more explanatory power on QI interventions than other healthcare improvement context models.

One potential weakness of MUSIQ may be that it does not try to distinguish the temporal ‘domains’ of the intervention (Walshe, 2007). These domains refer to the progress of the QI activity over time e.g. the organisational preconditions prior to commencing activity, the specific intervention actions and then sustaining the changes after an intervention. Including these domains would allow for better alignment and comparison with the business excellence frameworks. The contextual model proposed by Andersen et al. (2014) attempts this alignment by integrating Walshe’s intervention domains with Shortell’s dimensions of capability (Shortell, 1998, as cited in Andersen et al., 2014), to create a two dimensional framework of contextual factors. As well as considering the intervention lifecycle, the ‘capability dimensions’ arrange contextual factors into related groupings of cultural, technical, strategic and structural factors. These groupings may allow for more effective targeting and adaptation of solutions from other industry sectors or social science theory. This is undoubtedly useful; however Andersen et al.’s (2014) model does not attempt to show how the factors interact, beyond the simple placement of each factor at the relevant stage of the intervention lifecycle.

Table 2.1 compares the two models discussed. Despite the differences in the development and representation of these two models, it is clear that many important contextual factors (described as facilitators by Andersen et al.) are consistent. This is an encouraging finding in that it provides confidence that key contextual factors are now commonly identified by researchers (even though the precise definitions and labels are not yet consistent). It also supports the proposition that the appropriate next stage of research is focussing on how these factors actually interact.

Table 2.1: Comparison of the Andersen et al. (2014) model and MUSIQ

Change facilitators **Andersen et al. Model (2014)**	**Contextual factors** **MUSIQ, Kaplan et al. (2012)**
Management and leadership support	QI leadership - organisation level
Vision	
System-wide scope	Maturity of organisation QI programme; Task strategic importance to organisation

Change facilitators **Andersen et al. Model (2014)**	**Contextual factors** **MUSIQ, Kaplan et al. (2012)**
Customer focus	External motivators; Triggering events
Continuous improvement	Workforce focus on QI;
External support	Workforce focus on QI;
Measurement	Workforce focus on QI;
Supportive culture	Senior leader project sponsor; QI culture - organisation level; QI leadership - microsystem level
Holistic approach	Motivation for change; QI culture - microsystem level; Prior QI experience
Belief	Motivation for change; QI culture - microsystem level; Prior QI experience
Experience	Motivation for change; QI culture - microsystem level; Prior QI experience
Administrative support; IT-systems	Resource availability
Competence	Capability for improvement; Subject matter expert
Adaption - local translation of methods	Prior QI experience
Communication Alignment	QI (project) team diversity
Staff Involvement	QI (project) team diversity
Resources; accurate data	Resource availability
Physicians	Physician involvement; Physician payment structure
Teamwork	Team leadership; Team norms; Team decision making process; Team tenure
Training	Team QI skill

Potential advantages of MUSIQ include the theorised causal relationships between the interventional factor(s) (cause constructs) and the effect (improvement), and the positioning of the contextual factors within external, organisational, improvement team, and microsystem levels. The structure of MUSIQ is consistent with the *consolidated framework for implementation research* (CFIR) developed by Damschroder et al. (2009). The purpose of CFIR is to provide a consistent typology of domains and characteristics for conducting research into healthcare interventions (Damschroder et al., 2009).

Hovlid and Bukve (2014) provide an independent, recent example of the use of MUSIQ within their research into healthcare quality improvement. They report that MUSIQ's general hierarchy of influence from the organisation to the microsystem level was supported in their case study, but they also highlight additional potential weaknesses in the model, including the limited attention given to communication networks. This is a very useful observation, as communication is identified as essential to effective teamwork (Hoegl, & Gemuenden, 2001) and yet how people communicate seems to be somewhat buried in both models. Hovlid and Bukve also report that the boundary between the intervention and context was sometimes difficult to distinguish in practice, echoing the concern of Øvretveit (2010) noted previously. Introducing a new information technology system for example, could be an enabling organisational element (i.e. context) or the actual intervention. Finally, Hovlid and Bukve did not attempt to measure levels of any of the contextual factors, a consistently noted gap in the research (Hovlid & Bukve, 2014).

Regardless of the model being analysed, researchers seem to be agreed that the use of a model to explain (via hypotheses) how and why quality improvements succeed or fail is needed. Furthermore, a framework that can be generalised within a specific boundary or context is needed to increase the external validity of results (Damschroder et al., 2009; Portela et al., 2015). Understanding how contextual factors influence intervention outcomes is the appropriate next step in developing context-sensitive theories of healthcare quality Improvement. Testing the models via empirical data will then identify limitations and gaps and lead to their ongoing refinement (Andersen et al., 2014; Kaplan et al., 2013; Øvretveit, 2010).

2.5 Knowledge Gap and Research Questions

The most critical shortcoming of MUSIQ and equivalent models is in the handling of contextual factors in model testing. Although Andersen et al. (2014) and Kaplan et al. (2013) acknowledge that contextual factors are independent of the causal factors, in model testing, they have treated contextual factors as mediating factors (hence not independent of the cause factors) rather than moderating factors.

A mediating variable (say variable C) is a variable that is used to elaborate the nature of the relationship between the cause (say variable A) and the effect (say variable B). Thus for mediation to occur, variable A should be related to variable C, and variable C should be related to variable B (Baron & Kenny, 1996; Collins, Graham, & Flaherty, 1998). Thus if contextual factors are factors independent of the cause and effect, they cannot be treated as mediator variables.

A moderating variable is a variable that is independent of the cause (and effect), yet increases or decreases the strength of the relationship between the cause and the effect (Baron & Kenny, 1986). Thus it is clear that contextual factors should be handled as moderators rather than mediators. It is noted that moderation effects manifest as two-way interactions between the cause (in this study the Lean sociotechnical system) and the moderator (in this study the contextual factors) in explaining the effect (in this study the 'Effectiveness/Success') (Baron & Kenny, 1986; Hair, Hult, Ringle, & Sarstedt, 2014).

Having examined the literature on Lean in healthcare and contextual models for QI improvement in healthcare, including the recognised knowledge gaps, the research questions proposed for this study were:

RQ1: What are the specific contextual factors that influence the success of quality improvement initiatives within a New Zealand secondary public healthcare context?

RQ2: To what extent do specific contextual factors enable or inhibit the success of quality improvement initiatives within a New Zealand secondary public healthcare context?

2.6 Conclusion

The literature review reveals that QI in general, and Lean in particular, are active areas of focus with respect to healthcare. A call for consistent analysis frameworks is prominent, along with recognition that QI activity takes place within unique organisational contexts. Defining contextual factors properly is still required to allow greater transmission of effective strategies and methods internationally. A number of contextual models have been proposed; the model developed by Kaplan et al. (2012) is being actively considered by researchers. The contextual models would be enhanced if they can be used to examine and measure causal relationships. This knowledge gap has generated the research questions for the present study. Identifying contextual factors of interest (RQ1) is a limited contribution to the research without an accompanying attempt to measure cause-effect relationships and magnitude of influence (RQ2).

The next chapter outlines development of the theoretical model to explore the knowledge gap and answer the research questions.

Chapter Three

DEVELOPMENT OF THE THEORETICAL MODEL

3.1 Introduction

This chapter introduces the MUSIQ model as the starting point of the theoretical model. Section 3.3 then describes the process followed and supporting reasoning for bringing together MUSIQ and the Toyota Way. Section 3.4 describes the causal model, where the retained contextual factors are modelled as latent variables in a cause-effect relationship[1]. From this model, six hypotheses are generated for testing.

3.2 MUSIQ as the Initial Framework

The exploratory analysis of MUSIQ conducted by Kaplan et al. showed an apparent influence of the factors operating at the microsystem level (Kaplan et al. 2013). This finding is consistent with the model's implied hierarchy — the influence of the higher level factors such as resource availability, improvement culture or organisational leadership, is presumably condensed through each level, and it is reasonable to assume that there will be a stronger *measurable* relationship between variables and outcomes closer to the actual process changes. Another way of describing this is that there is a concentrating effect as one proceeds down through the hierarchy layers of the model, and it is harder to measure causal influence (as well as control for other factors) between variables at the higher levels of the model (such as organisational leadership) and any specific process improvement outcome.

The first step for the present research was to develop a causal model, based on the literature, to define and test the variables at this microsystem level of MUSIQ. The causal model allows quantitative modelling of survey data to test the hypotheses; some of these hypotheses involve transitionary contextual factors, as per Worchel's (1986) definition, which enable the researcher to examine how these factors might influence the process activity that leads to outcome improvements. The MUSIQ contextual factors (Kaplan et al., 2012) selected as exogenous variables were *motivation, QI capability, team norms, decision making and system and process changes* (refer Figure 2.1).

[1] To assist the reader distinguish between the labels of the latent variables as precisely defined for the model from any general usage of common words such as motivation or teamwork, specific references to the latent variables of the model are capitalised and enclosed within single speech marks in the text.

3.3 Isolating and Operationalising the Cause(s)

This study focused on healthcare practitioners who had incorporated Lean into their improvement activity. The challenge is to represent the Lean sociotechnical system as a cause construct. As a theoretical foundation, the Toyota Way (TW) model was used to provide a potential explanation for the 'system and process changes', and also to offer an alternative understanding of the *QI capability* contextual factor in MUSIQ. The TW is the framework of principles and values underpinning the Toyota production system (Jayamaha, Wagner, Grigg, Campbell-Allen, & Harvie, 2014). It is represented as two pillars: "Continuous Improvement" and "Respect for People" (Figure 3.1). The pillars are in turn comprised of five elements: — challenge (taking a long-term perspective to improvement actions, meeting challenges with courage), kaizen (incremental improvement actions themselves), genchi genbutsu (go and see), respect and teamwork (Jayamaha, et al., 2014).

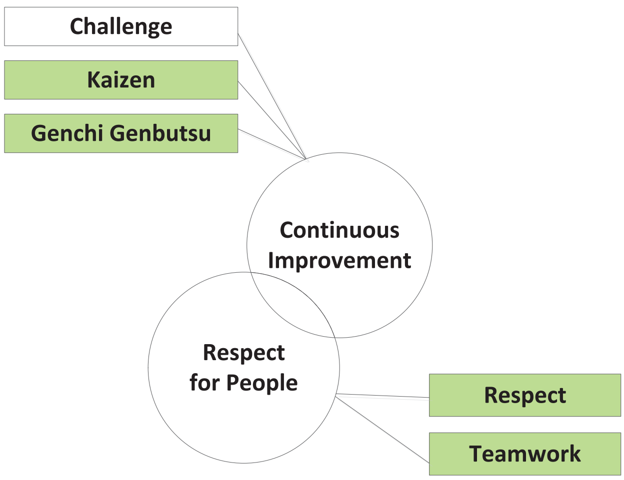


Figure 3.1: The Toyota Way Conceptual Framework
(Toyota Motor Corporation, 2001).

3.3.1 Respect and Teamwork

In the Toyota Way context, 'Teamwork' (Figure 3.1) sets the platform for 'Respect for People' to come into fruition (Ichijo & Kohlbacher, 2008; Liker & Hoseus, 2009). Hoegl and Gemuenden (2001) define teamwork as the quality of *interactions* within teams, as distinct from the activities themselves and the effectiveness of those activities and tasks (task skill). Focusing on the interactions also defines a boundary with other related and potentially confounding team factors such as team diversity, experience or leadership. ‘Teamwork’ in this model is analogous to the contextual factor *Team Norms* in MUSIQ.

‘Respect for People’ reflects the personal behaviours and feelings within a team (e.g. “enjoyment”), as distinguished from functional or transactional interactions (e.g. the frequency of meetings). Hoegl and Gemuenden (2001) describe these elements of behaviour as ‘sentiments’ which *“influence the interactions and activities, and are in turn, influenced by them”* (Hoegl & Gemuenden, 2001, p.346). The Shingo model offers a slightly different perspective, concentrating on the behaviours arising from shared fundamental principles to provide the cultural foundation for an organisation (Machado Guimarães & Crespo de Carvalho, 2014). However, the personal feelings of workers within a team still presumably influence their interactions, regardless of shared principles. Thus any equivalence between sentiment and principle (when the terms are used in this manner) is not exact. Because this study evaluated Lean as the primary improvement methodology, ‘Respect for People’ was proposed as the label for this behaviour/sentiment variable. The ‘Respect for People’ pillar elements of the TW relate to the behaviour and culture of the organisation (‘Human Resource Capability’), which is the *intangible* (and inimitable) resource capability of Toyota’s sociotechnical system (Jayamaha et al., 2014; Rother, 2010).

3.3.2 Continuous Improvement

The elements under the Continuous Improvement (CI) pillar of the TW (Figure 3.1) relate to the specific QI activities and tasks which Jayamaha et al. (2014) refer to as the *tangible* resource of Toyota’s sociotechnical system. The TW framework clearly indicates that CI is a broader concept than kaizen alone, with additional strategic and behavioural dimensions.

3.3.3 Integrating TW and MUSIQ

Table 3.1 reconciles the elements of the model used in this study against the elements of the TW and MUSIQ. There is strong alignment between the TW elements and the MUSIQ elements, and thus a plausible theoretical basis to consider Lean as an explanatory mechanism within the MUSIQ concept of ‘system and process changes’. MUSIQ is a generic model (MUSIQ has not been specifically designed for a Lean sociotechnical system) that outlines the variables that stand in the way between actions (QI interventions) and results (Improvement Outcomes). These transitory variables have been labelled as contextual variables. Applying the TW extends MUSIQ by considering the cultural elements of a Lean sociotechnical system, alongside the existing MUSIQ team norms and capability factors.

Table 3.1: Alignment of the MUSIQ and TW Models Used to Develop the Causal Model

MUSIQ Contextual Factor(s)	**Toyota Way Element**	**Latent Variable in the Causal Model**
• System and Process Changes†	• Kaizen • Genchi Genbutsu	• 'Lean Actions'†
• Team Norms, Skill and Decision Making* • QI Capability*	• Teamwork • Respect	• 'Teamwork'†
• Team Norms*	• Respect	• 'Respect for People'†
• Motivation*		• 'Motivation'*
• Improvement OutcomesΔ		• 'Perceived Success'Δ
† The cause variables; Δ The effect (outcome) variable; * The contextual variables (moderator variable) that was tested		

3.4 Incorporating the Contextual Factors

The hypothesised causal model, in readiness for empirical testing, is shown in Figure 3.2. Each element of the model and their interrelationships are now described.

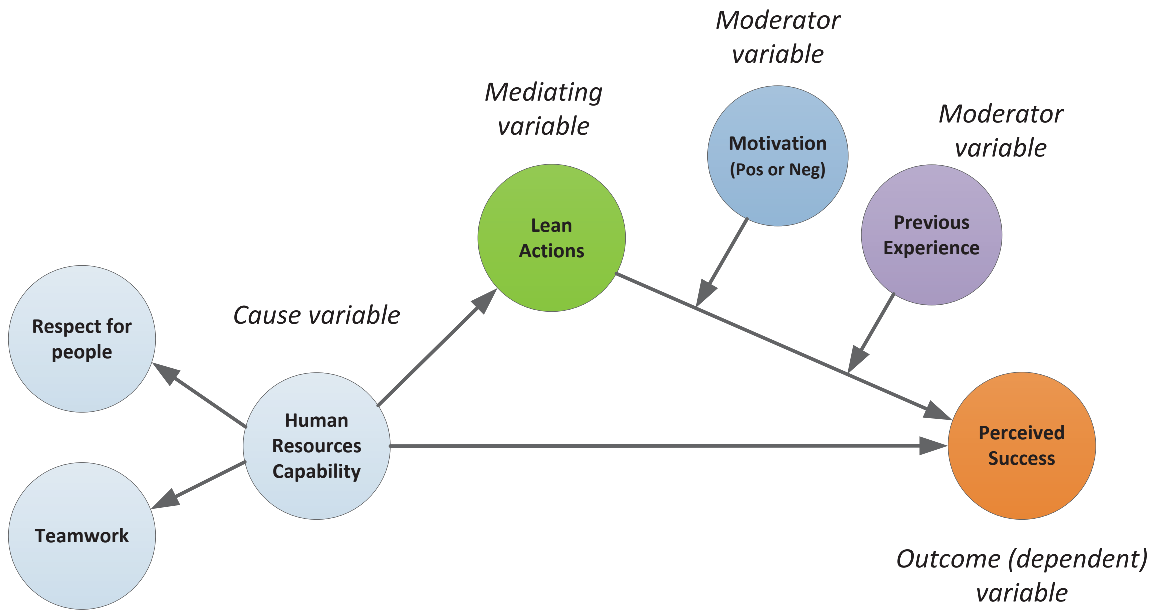


Figure 3.2: Hypothesised causal model

The ‘Teamwork’ and ‘Respect for People’ variables were combined into a more abstract (second order) construct to represent the ‘Human Resources (HR) Capability’ of the Lean sociotechnical system. As argued by a number of researchers (Emiliani, 2006; Jayamaha et al., 2014; Ohno, 1988; Rother, 2010), it is the ‘HR Capability’ of the Lean Management System that causes ‘Lean Actions’ to produce results (i.e. ‘Human Resources Capability’ → ‘Lean Actions’ → Success). Since the results were indirectly captured as perceived by survey respondents, Success was labelled as ‘Perceived Success’ (Figure 3.2). It is also important to note that MUSIQ’s contextual variable ‘QI Capability’ is subsumed in the ‘Human Resource Capability’ construct (i.e. a more capable human resource system, amongst other attributes, has a higher QI capability).

‘Motivation’ and ‘Previous Experience’ were treated as moderators affecting ‘Lean Actions’ — enhancing or attenuating the activity creating the changes. In this regard, these moderators are also the contextual factors affecting Lean action. In keeping with the definition used by Worchel (1986), a contextual factor was defined as a *transitionary factor* that interacts with the cause (in this case ‘Lean Actions’) to change the outcome (in this case ‘Perceived Success’). To avoid any confusion with the previous paragraph, it is important to note that, as the mediator variable, ‘Lean Actions’ was hypothesised as the mechanism *through which* ‘HR Capability’ achieves results. It is in this mediator role that ‘Lean Actions’ are functioning as the cause. ‘Motivation’ was hypothesised as an amplifier or attenuator on the activity (‘Lean Actions’); this is because, depending on the motivators, some motivators (e.g. belief in the end goal) can have a positive interaction effect on ‘Lean Actions’ while other motivators (e.g. lack of management support) can have a negative interaction effect on Lean Action (Damij, Levnajić, Skrt, & Suklan, 2015; Hackman & Oldham, 1976; Seo, Barrett, & Bartunek, 2004; Stajkovic & Luthans, 2001). An employee (or associate, in the language of Lean/quality management) is said to be positively motivated when he or she experiences “meaningfulness” in their work, which eventually leads to greater performance outcomes (May, Gilson, & Harter, 2004; Hackman, & Oldham, 1976).

The constructs for motivation were based on the goals and rationale for the improvement activity. Self-determination theory (Gagné & Deci, 2005), articulates motivation as a continuum from amotivation, through to extrinsic motivation and finally on to intrinsic motivation. Implicit in the concept of high performing teams is a high level of intrinsic motivation. Comparing self-determination theory with the earlier goal setting theory of Lock and Latham (Locke and Latham, 1990; Gagné and Deci, 2005), Gagné and Deci note the convergence between a “meaningful rationale” facilitating internalisation in self-determination theory and goal ‘acceptance’ in goal-setting theory (Gagné and Deci, 2005, p.341). The distinction between positive and negative motivation is determined by the type of psychological and emotional impact the contextual factor has on the individual carrying out a task (Seo, Barrett, & Bartunek, 2004). Seo et al. developed a two dimensional affective experience structure spanning pleasant to unpleasant (x axis) and activation to deactivation (y axis). Different emotional states exist at various points along these axes and influence the strength and the *direction* of associated behaviour. Positive motivation indicators represent an activated and/or pleasant psychological environment for the associates (workers) — an environment in which associates will feel good about themselves because of management support for the QI initiatives that they engage in, experience meaningfulness of the tasks (knowing why these are being done), and perceiving that the time and effort of the people have been well utilised. Negative motivation indicators on the other hand represent a deactivated and/or unpleasant psychological environment — a perception of unrealistic goals and feeling that their effort is not effective as there are better ways to do/approach the work. Negative motivation, resulting in deactivation, will result in diminished effort or emotional disengagement (Seo et al., 2004).

Finally, in keeping with the MUSIQ framework, ‘Previous Experience’ or training in QI was added as another potential moderator having a positive interaction effect on ‘Lean Actions’. ‘Previous Experience’ was hypothesised as a positive moderator because when the ‘Previous Experience’ is high, for the same Lean Action, the organisation should achieve a greater success (based on a learning curve effect) than when the ‘Previous Experience’ is low.

Six hypotheses were generated for testing using the causal model:

> ***H1:*** *‘HR Capability’ has a positive effect on ‘Perceived Success’;*
>
> ***H2:*** *‘HR Capability’ has a positive effect on ‘Lean Actions’;*
>
> ***H3:*** *‘Lean Actions’ have a positive effect on ‘Perceived Success’;*
>
> ***H4:*** *‘Previous Experience’ has a positive moderating effect on ‘Lean Actions’;*
>
> ***H5:*** *‘Positive Motivation’ has a positive moderating effect on ‘Lean Actions’ in improving ‘Perceived Success’;*
>
> ***H6:*** *‘Negative Motivation’ has a negative moderating effect on ‘Lean Actions’ in improving ‘Perceived Success’.*

3.5 Conclusion

The MUSIQ model does not specify any QI methodology within the generic label of "system and process changes" Given that Lean or methods based on Lean are increasingly part of the QI picture in healthcare, at Least in New Zealand, Australia, UK and other countries whose health systems have strong links to the NHS tradition, it is reasonable to assume MUSIQ should be applicable to Lean activity. The researcher first sought a plausible theoretical basis to confirm the compatibility of MUSIQ and Lean. The Toyota Way was proposed as the fundamental framework that enables subsequent Lean activity, and a parallel lens which can be used to consider contextual factors. Compared alongside the subset of MUSIQ microsystem factors considered in this study, there is a high degree of alignment. The contextual factors were then modelled in a cause-effect relationship to specify the role of each latent variable as cause, mediator, moderator or outcome variable, in readiness for empirical testing via multiple regression statistical modelling. The next chapter outlines the methodology adopted in addressing the research problem/research questions.

Chapter Four
METHODOLOGY

4.1 Introduction

This chapter begins with a consideration of the research strategy decisions, including the background to the decision to adopt a positivist paradigm (sections 4.2 and 4.3). Section 4.4 discusses the suitable positivist data analytic techniques to test the hypotheses, including justification of the selection of the partial least squares path modelling (PLSPM) approach. Section 4.5 details the ethical considerations followed and Section 4.6 covers the data collection strategy including survey population and sampling. Section 4.7 provides a detailed description of the selection of measurement items for each of the latent variables in the causal model. The chapter concludes with a description of the survey pre-testing and delivery.

4.2 Research Paradigm

4.2.1 Defining a Research Paradigm

A paradigm is a certain belief system that guides the researcher in the way in which he/she conducts the research (Guba & Lincoln, 1994). One important aspect associated with a paradigm is the question what is reality? This is answered by the ontology associated with the researcher's paradigm (Bryman, 2012; Guba, 1990). The second aspect associated with a paradigm is the question how does one know something? This is answered by the epistemology associated with the researcher's paradigm (Bryman, 2012; Guba, 1990). The third important aspect that naturally follows from the epistemology is the question how does one go about in finding out something? This is what is known as the researcher's methodology (Creswell & Poth, 2017; Guba, 1990, Yilmaz, 2013).

Although alternative paradigms exist in social research, the two dominant mainstream paradigms that compete with one another — because they are polar opposite, both ontologically and epistemologically — are positivism and constructivism, also known as interpretivism (Bryman, 2012; Guba, 1990).

The positivistic ontology holds that reality exists irrespective of the observer (i.e. reality is singular and objective). The positivistic epistemology therefore holds that knowledge claims are best made by formulating and testing cause-effect hypotheses by way of making precise measurements of the variables associated with researchers' hypotheses. Thus positivistic methodology is often associated with quantitative research involving statistical hypothesis testing (Bryman, 2012; Creswell, 2002). A variant of positivism is post positivism; although post positivists believe in the positivistic ontology and epistemology, they believe that triangulation, in terms of data and/or methods can enhance one's understanding of the imperfect and complex social world (Crewell, 2002; Zammito, 2004). Both positivists and post positivists emphasise what is often labelled as "rigour" (or "scientific rigour"), which refers to the reliability and validity of the data and the concepts being measured, as they are central to statistical conclusions validity the overall validity of statistical conclusions (Campbell, Parks& Wells, 2015; Malterud, 2001).

The interpretivist (constructivist) ontology holds that reality is subjective and multiple. Thus the interpretivist epistemology holds that knowledge is socially constructed and therefore the context, as well as the people/informants who are associated with the phenomenon being studied, becomes part of the reality. Thus interpretivists tend to rely on qualitative data and the trustworthiness of such data/information being provided by the informants (Creswell & Poth, 2017).

4.2.2 Ontological and Epistemological Considerations

The rationale for adopting a positivist epistemological approach over an interpretivist stance must first be outlined. Primarily this decision was based on the observation that theories (in this case theories of how contextual variables affect QI), however rudimentary they might be, have already been proposed. From this starting point, the next logical step was deemed to be the testing of these theories by generating hypotheses and empirically testing them – the deductive approach to building knowledge (Bryman, 2012). Supporting this decision, one of the most important themes in the literature on Lean and other QI methods in Healthcare is the lack of empirical research, the norm for physical sciences, which clinical medicine certainly follows. Although healthcare service delivery and health quality management are not necessarily the same as evidenced-based, clinical medicine itself, solid empirical support for any potentially disruptive intervention is a *sine qua non* for the clinical leaders managing the overall health system.

Countering the positivist assumption that healthcare QI should produce objective, observable facts that can be quantified and measured, it would be difficult to describe a situation where the ontological position of social constructivism could be better applied than to the concept of 'Teamwork' within an organisation. Constructivism posits that social phenomena (such as the teamwork, team norms, and interactions considered in this study) are continually 'constructed' by the social actors involved (Bryman 2012). Intuitively, any emergent phenomena produced by these actors must be a product of those social actors and will therefore be different if different actors are involved. The role of individuals in "creating social reality" (Bryman, 2012, p. 34) tends to lead on to qualitative research strategies that emphasise understanding as much as possible of research participants' individual perspectives i.e. each setting is variable, and subjective decisions are inevitably made by the researcher as to what features of any given social research setting are emphasised (Bryman, 2012). In terms of research design, case studies are common in qualitative research, with focus groups, observations and interviews common research methods. Results from different studies and even actors within the same study may not be able to be generalised to a larger group or different context. Qualitative designs are common in much of the non-clinical healthcare QI research (Andersen et al., 2014). As noted by Øvretveit (2010), and Walshe (2007), amongst others, this dominance of qualitative research limits replicability, potentially limiting the understanding and acceptance of Lean in healthcare.

Taking all the above into account, the researcher decided that for the present study a positivist epistemological framework of hypothesis testing, with an objective ontological orientation, was narrowly preferred over an inductive, theory building epistemological approach and constructivist ontological perspective. The quantitative research design outlined below followed naturally from this decision. This does not imply that qualitative research has no further place in the study of this topic, and important knowledge gaps in the understanding of teamwork social interaction phenomena are identified and discussed further in chapter 7.

4.3 Research Design Strategy

It is considered rare that a social research theory or model is generated entirely without antecedents from previous work (Dubin, 1978). An alternative strategy to starting from scratch is to refine and improve existing models by empirical testing. As previously noted, the requirement to test context-sensitive models via empirical data was a prominent theme in the literature review, and this approach will then identify limitations and gaps in these models and lead to their refinement (Andersen et al., 2014; Kaplan et al., 2013; Øvretveit, 2010).

A survey consisting of multiple item scales was considered the appropriate strategy, given the variables of interest cannot be directly observed or measured (DeVellis, 2012). Portela et al. (2015) provide useful guidance specifically for studying improvement interventions in healthcare. They outline Weiss's *logic of analysis in evaluation*, which categorises the type of question being asked and the type of data being examined. The task of looking for combinations of actors, services and conditions that are associated with success and failure is described as a 'profiling' task, with surveys being a common and justified method used for profiling (Portela et al., 2015, p. 331-332).

Case studies remain a viable strategy for this research, but case studies are usually of more value in drilling down deeper into causal mechanisms to unravel rich context-bound information associated with a causal phenomenon (Philliber, Schwab & Sloss, 1980). As such, case studies are particularly suitable to answer research questions that begin with "why" or "how" (Yin, 2013). However, the literature review also indicated that broad preliminary research on QI aimed at formulating and testing universal theories examining the context is still needed. Therefore there is a need for working with larger samples and observations, before establishing where to dig deeper in search of context-bound information via case studies.

The research was not considered suitable for an experimental research design. There are too many variables (some are unknown) that constantly interact or confound with the causal and contextual variables and controlling such variables is practically impossible. This challenge was highlighted in the literature and the reason for the calls for a context-driven analysis in the first place (Walshe, 2007). It is also assumed to be impractical to waste scarce resources creating control groups which might realistically be expected to fail. Therefore the researcher adopted the survey research approach to test the theoretical model (causal model).

4.4 Possible Data Analytic Methods for Hypothesis Testing

The six hypotheses (section 3.4) involve seven latent variables (constructs) (Figure 3.2). Each latent variable is operationalised through multiple measures (the survey questionnaire items) to improve reliability and validity of the constructs and thereby, the accuracy of the parameter estimates, given the sample size.[2] The exogenous variable 'Human Resources Capability' is a latent variable that is high in abstraction (a second-order construct), which is reflected by two (first order) latent variables: 'Respect for People' and 'Teamwork'. Use of multiple measures to represent a construct requires a way of representing multiple dimensions as unidimensional measures (Hair et al., 1998). A construct can be represented as either a component or a common factor (Grigg & Jayamaha, 2014; Mulaik, 2009). Having selected this representation, the next step would be examine the relationships between the dependent variable and the interdependent variable(s) of the structural relationships representing the researcher's hypotheses (Grigg & Jayamaha, 2014). Given the above requirements, three data analytic methods were considered by the researcher: covariance structure analysis, principal components regression, and partial least squares path modelling (Grigg & Jayamaha, 2014). Each of these approaches is outlined briefly below.

4.4.1 Covariance Structure Analysis

The covariance structure analysis, which is also known as covariance based structural equation modelling (CBSEM), is routed in the common factor analysis approach (Grigg & Jayamaha, 2014; Kline, 2011). After specifying the linear structural relationships between a construct and its measures (i.e. the measurement model), and between constructs (i.e. the linear structural regression models), CBSEM calculates a global optimisation parameter. This optimisation parameter seeks to minimise the discrepancy between the *model-implied* covariances between the measures (i.e. covariances calculated from the estimated parameters) and the *observed* covariance between the measures (Grigg & Jayamaha, 2014; Kline, 2011). Since CBSEM optimises all the structural relationships using a single global optimisation parameter, CBSEM is also known as a full information structural equation modelling approach (all unknown parameters are estimated through a single optimisation procedure) (Kline, 2011). Thus

[2] The true parameter estimate (or the parameter estimate estimated from population data) will, in general, be different from parameter estimates derived from a sample due to sampling error and all things being the same, a smaller sample will provide less accurate parameter estimates due to higher standard error (Hair, Anderson, Tatham, & Black, 1998).

it is not surprising that CBSEM is *de rigueur* for many social and behavioral science disciplines and academic journals (Kline, 2011; Mulaik, 2009).

Although CBSEM is the most acceptable causal modelling approach in social and behavioural sciences, it is a parametric method, which is highly sensitive to violation of parametric assumptions. In addition, CBSEM is an asymptotic (large sample) method that does not perform well with small samples (Byrne, 2010; Iacobucci, 2010; Kline, 2011). The CEBSEM approach also encounters problems when the specified model involves complex measurement models such as formative constructs, large number of measurement items, and second-order constructs (Grigg & Jayamaha, 2014). Finally, CBSEM is also not encouraged in situations that involve new concepts/constructs or concepts whose meaning keeps evolving (Chin, 1998).

The primary reason for excluding CBSEM for the present study is data on measurement items failing to meet the parametric assumption of normality (i.e. normal distribution). Table 4.1 depicts some distributional characteristics (skewness and kurtosis) along with Anderson and Darling (AD) goodness-of-fit test results for Normality (Anderson & Darling, 1954) for the first three and last three survey items of the survey questionnaire[3]. The null hypothesis in the AD test is that data come from a normally distributed population and therefore a significant *p* value ($p < 0.05$) fails to retain this contention (Anderson & Darling, 1954). The information depicted in Table 4.1 clearly indicates the violation of normality.

Table 4.1: Evidence of Nonnormality – The First Three and Last Three Survey Items

Survey Item	Skewness	Kurtosis	Anderson-Darling Test For Normality		Normality Shown?	Comments
			A^2	*p*-value		
I1	-1.13	2.52	9.24	< 0.005	No	Data very negatively skewed and kurtotic
I2	-0.59	0.89	10.75	< 0.005	No	Data negatively skewed and kurtotic
I3	-1.10	1.70	7.27	< 0.005	No	Data very negatively skewed and kurtotic
I39	-0.85	1.54	8.35	< 0.005	No	Data negatively skewed and kurtotic

[3] The researcher is grateful to Dr Jayamaha for conducting the Normality tests

Survey Item	Skewness	Kurtosis	Anderson-Darling Test For Normality		Normality Shown?	Comments
I40	-0.19	-0.00	5.38	< 0.005	No	Data negatively skewed
I41	-1.24	1.99	2.56	< 0.005	No	Data very negatively skewed and kurtotic

4.4.2 Principal Components Regression

In principal components regression (PCR), each construct is represented as the first principal component of the measures that operationalise the construct. The scores of the first principal components are then used to estimate the parameters associated with the specified structural relationships between constructs using the multiple regression approach (Haenlein & Kaplan, 2004; Mevik & Wehrens, 2007). PCR is very flexible and any complex model can be tested using PCR (Haenlein & Kaplan, 2004; Mevik & Wehrens, 2007). However, since PCR is a limited information approach that is heavily weighted towards the measurement model (Grigg & Jayamaha, 2014), this approach was not considered for the present study.

4.4.2 Partial Least Squares Path Modelling

Partial least squares path modelling (PLSPM) is also a component-based limited information approach, but it occupies the middle ground between the rigour found in CBSEM and the flexibility found in PCR (Grigg & Jayamaha, 2014). Although PCR and PLSPM are very similar (often both approaches produce very similar parameter estimates as observed by Haenlein and Kaplan, 2004), the knowledge base of the latter is continuously being updated and more and more procedures are emerging in the PLSPM literature to mimic some of the features found in CBSEM (Hair el al., 2014).

Since PLSPM uses a nonparametric method (e.g. bootstrapping, jack-knifing) to determine the statistical significance of model parameters, PLSPM is not reliant on parametric assumptions (Grigg & Jayamaha, 2014; Hair el al., 2014). PLSPM was chosen as the structural equation modelling method for the present study primarily to overcome the problem of non-normal data (see section 4.4.1). This decision was supported by other reasons such as the newness of some of the constructs (e.g. 'Positive Motivation' and 'Negative Motivation'), model complications such as factor interactions (the moderator variables in the model by definition interact with the causal factors). The presence of a second order construct also contributed towards the selection of the PLSPM approach.

4.5 Ethical Considerations

The Massey University ethics guidelines (2015) were reviewed and the risk screening questionnaire completed (Massey University, 2015). Ethical considerations for this project were discussed with CDHB staff, management, and potential survey participants. The fundamental ethical criteria of *voluntary participation, informed consent, maintenance of privacy and confidentiality, respect for participants and sharing of results* (Burns, 2000), were all relevant to this project and were addressed via the communication plan and management of the survey. No patient identification or clinical information was accessed or analysed in this research. Potential conflict of interest or analytical bias was avoided by excluding any projects which the researcher had participated in from the sample. To encourage open and frank responses from participants, confidentiality was emphasised alongside neutrality of outcomes (i.e. respondents were encouraged to say something did not work as planned if that was the case).

4.5 Data Collection

A survey instrument was developed to operationalise the seven constructs of the causal model, for the purpose of collecting data to test the hypotheses. Each survey question that operationalised a construct contained a statement for which agreement was sought using a 5 point Likert scale ranging from strongly disagree (coded 1) through to strongly agree (coded 5); more details follow. The survey instrument also included 6 open ended survey questions to assess respondents' exposure to QI and understand the metrics that have been used in measuring results (success).

4.5.1 Survey Population

Survey participants were selected from three separate QI programmes operating within the Canterbury district health board (CDHB), one of the largest DHBs in New Zealand. The three QI programmes are given below.

1. A ward-based programme including Lean thinking and methods, based on the UK National Health Service *Productive Ward* model — branded internally as "Releasing time to care";

2. An internally developed, two-day staff training programme called "Collabor8", where a small-scale practical improvement project is undertaken and then considered in terms of Lean principles, change management and quality assurance fundamentals;
3. Independent quality improvement projects undertaken by quality facilitators and miscellaneous clinical and non-clinical staff.

Treating all three programmes as the initial sampling frame provided a sufficient quantity and a diverse range of Lean/QI initiative sizes and intervention types — including clinical projects such as changes to models of care, as well as non-clinical operational improvements to service delivery. The criteria for inclusion were:

1. The improvement activity fits the definition of *"purposeful efforts to secure positive change"*.
2. The improvement activity had been undertaken within the last 24 months. This time period was chosen so that the outcome could be meaningfully assessed, the project was still relatively fresh in respondents' memories, and there was a sufficient pool of initiatives to draw from.

4.5.2 Sampling Methodology

The releasing time to care participants were surveyed in two separate cohorts based on their location, and the Collabor8 and independent projects were surveyed together in a third cohort. The survey population was treated as a single frame for analysis, and as a nonprobability sample (Bryman, 2012). All of the population with valid email addresses was invited to respond. The issue of respondent bias which arises from this approach is examined in the discussion of results.

4.6 Survey Construction

The contextual factors as described in the MUSIQ model are very high-level concepts, each capable of being broken down into a more detailed taxonomy. Given the potential for an overwhelming number of sub elements, the aim of the survey construction was to develop a minimum set of survey items that would still provide for acceptable content validity and measurement reliability for each latent variable in the model.

Alongside the previously noted suggested influence of factors operating at the microsystem level of MUSIQ, some means of scaling MUSIQ down was also considered necessary for purely practical reasons. Attempting to measure the entire MUSIQ framework using multi-scale survey items for good reliability would make the resulting survey highly impractical to administer and complete. In their exploratory testing Kaplan et al. acknowledge the use of single measurement indicators for many items and they still ended up with a survey containing 70 questions (Kaplan et al., 2013). Survey professionals suggest diminishing attention/concentration spans adversely affect longer surveys (Mora, 2011); this was an additional driver to limit the survey. The decision to restrict to the microsystem level was also a good fit with the proposed survey population, who predominantly worked at this level.

4.6.1 Measurement Items that Operationalise 'Lean Actions'

Given the breadth and depth of the concepts, methods and tools that might be described as "Lean", the objective was not to create an exhaustive list of Lean items, but a condensed set of survey items that would constitute a reliable indicator of Lean Activity when treated as a Likert measurement scale. Beginning with the dimensions of Lean thinking as articulated by Womack and Jones (2003), questions were developed to assess the extent Lean concepts including *waste, value, flow of value, standardisation and continuous improvement* contributed to the initiative. Complementing the Lean thinking concepts, the Lean Body of Knowledge as taught by the American Society of Manufacturing Engineers provided a useful reference as to a practical minimum set of methods and tools typically identified as part of the Lean 'Toolkit' — visual management, standard work, root cause analysis, genchi genbutsu (observations at the place of work), and PDSA cycles (American Society of Manufacturing Engineers, 2008). This balance of Lean concepts and tools resulted in the following condensed set of survey items examining Lean activity as set out in Table 4.2.

Table 4.2: The Survey Items Included in 'Lean Actions'

Latent Variable	**Survey Item**
'Lean Actions'	I1. Our initiative identified waste in work processes.
	I2. Our initiative identified value-adding activity in work processes.
	I3. Our initiative attempted to identify the underlying causes of process problems.
	I4. Our initiative observed operational staff in their workplace.
	I5. Our initiative used visual tools as part of the solution to support operational processes.
	I6. Our initiative attempted to develop safe, reliable and efficient procedures for staff to follow
	I7. Our initiative used a Plan-Do-Study-Act framework.

For the other variables within the model, the literature review indicated that there might be existing survey instruments which could be repurposed for this research, and that this would be the preferred strategy for reducing unnecessary effort and duplication. Using existing surveys also helps maintain consistency with previous research and further validates well developed but infrequently used survey instruments. This goal was achieved for 'Teamwork' and 'Respect for People' but not the remaining constructs.

4.6.2 Measurement Items that Operationalise 'Teamwork' and 'Respect for People'

Kaplan et al. (2012) developed questions for measuring team decision making and behaviour norms from the QI practices index developed by Lemieux-Charles et al., (2002). This study also started with the work of Lemieux-Charles et al., (2002). 'Teamwork' questions were further compared to the survey developed by Schouten, Grol, and Hulscher, (2010), and the 'Teamwork Quality' theoretical concept, as articulated by Hoegl and Gemuenden (2001). The Hoegl and Gemuenden framework of teamwork quality is made up of six elements: *communication, coordination, balance of member contributions, mutual support, effort and cohesion* (Hoegl and Gemuenden 2001). These elements served as the foundation for content validity assessment.

Although there are some minor differences across these three sources when categorising contextual factors and improvement actions, there is also a high degree of convergence between them. This provided confidence in the content validity of the ‘Teamwork’ and ‘Respect for People’ variables in the model.

A detailed systematic review of potential instruments, relevant to healthcare improvement, is provided by Brennan, Bosch, Buchan, & Green, (2012), although the emphasis is on primary care. The work of Brennan et al., (2012) is invaluable to understand the theoretical basis, intended use and previous testing of these instruments. Unfortunately, as they acknowledge, many of these surveys are limited by definitions of constructs that are hard to re-use, or else require further validation (Brennan et al., 2012, p. 13). Examining the content and purpose of these instruments as summarised by Brennan et al. did not reveal any further usable examples to adapt. To complete the remainder of the hypothetical model, new items for ‘Perceived Success’, ‘Motivation’, and ‘Previous Experience’ were developed.

4.6.3 Measurement Items that Operationalise ‘Perceived Success’

Prior to constructing the survey, a preliminary review of initiative outcomes amongst the survey sampling frame revealed a poor level of measurement standardisation. Therefore the approach adopted by Kaplan et al. (2012), to represent respondents’ *perception* of the success of their initiatives was used as a proxy measure. This measurement strategy was justified by the responses. Looking ahead briefly to the results — although 48 of 105 respondents reported the existence of measurable data associated with their initiatives, when examined this data was not able to be collated, combined or compared in any meaningful way. An objective measure of success was essential for the study, but having a single measurement source did introduce the risk of common method error into the results.

4.6.4 Measurement Items that Operationalise ‘Motivation’

Although the motivation of individuals within organisations is a heavily researched topic, surveys relating specifically to the motivation of individual healthcare practitioners in relation to quality improvement activity proved elusive. New survey items were developed after referring to a range of current theories of QI motivation in the healthcare context. Dolea and Adams (2005) provided a theoretical overview of the established theories on motivation and an examination of the empirical evidence supporting these theories in the healthcare setting. Measurement items for motivation were developed based on the theory of a “meaningful rationale” for work to ensure intrinsic motivation (Gagné & Deci, 2005). May et al. (2004) define *experienced meaningfulness of work* as “value of a work goal or purpose, judged in relation to an individual’s own ideas or standards” (p. 41).

Positive motivation indicators represent a ‘safe’ psychological environment for the associates (workers) — an environment in which associates will feel good about themselves because of management support for the QI initiatives that they engage in, experience meaningfulness of the QI initiatives (knowing why these are being done), and perceiving that the time and effort of the people have been well utilised for the QI initiatives. Negative motivation indicators on the other hand represent a ‘discouraged’ psychological environment for the associates, where they feel management has set them unrealistic goals, is unsupportive, or that their effort was not effective as there are better ways to do/approach the QI initiatives.

Note that this operationalisation of negative motivation does not include external influences which may act as a driver or inducement to engage in the improvement activity for inappropriate reasons. Examples of this kind of negative motivation might be acting because of coercion or fear of an undesirable outcome if an instruction is not complied with (Linder, 1998). To prevent any confusion, ‘Negative Motivation’ as defined here refers to the emotional state of an associate, i.e. an individual negative affective state as per Seo et al. (2004), in turn reducing the sense of meaningfulness and engagement to contribute (May et al. 2004). Two aspects of QI work that reflected this aspect were chosen: perceiving that a chosen solution was not the optimum way of carrying out the task and perceiving that the tasks or goals imposed were unrealistic. Table 4.3 displays all the survey items as used for the quantitative analysis, mapped to the appropriate latent variable/construct. A complete list of the survey items is also attached as Appendix 1. The complete survey list includes those items removed during reliability testing and Principal Components Analysis (PCA), as well as the additional questions asked covering respondent demographics and initiative details. These questions were included to allow for subgroup stratification and internal CDHB consideration of project types and locations.

Table 4.3: Survey Items Used for Quantitative Analysis

Items Retained after Reliability and Principal Components Analysis	**Mapping to Latent Variables**
I1. Our initiative identified waste in work processes.	'Lean Actions'
I2. Our initiative identified value-adding activity in work processes.	'Lean Actions'
I3. Our initiative attempted to identify the underlying causes of process problems.	'Lean Actions'
I4. Our initiative observed operational staff in their workplace.	'Lean Actions'
I5. Our initiative used visual tools as part of the solution to support operational processes.	'Lean Actions'
I6. Our initiative attempted to develop safe, reliable and efficient procedures for staff to follow	'Lean Actions'
I7. Our initiative used a Plan-Do-Study-Act framework.	'Lean Actions'
I9. We maintained a focused effort.	'Teamwork'
I11. Team members took responsibility for their actions and behaviours.	'Respect for People'
I12. We avoided personal criticism when reviewing ideas within the team.	'Respect for People'
I14. I enjoyed being part of this team.	'Respect for People'
I15. Team members supported each other.	'Respect for People'
I16. We had clear goals.	'Teamwork'
I17. We had regular meetings.	'Teamwork'
I18. We used facts and data to understand performance.	'Teamwork'
I19. We knew how we were progressing towards our goals.	'Teamwork'
I20. All of our team were encouraged to contribute ideas.	'Teamwork'
I21. I believed this initiative was an appropriate use of our time and effort.	Motivation 1 (positive indicator)

Items Retained after Reliability and Principal Components Analysis	**Mapping to Latent Variables**
I22. I believed there were better ways to improve the focus area than the solution we chose.	Motivation 2 (negative indicator)
I27. Our goals were unrealistic.	Motivation 2 (negative indicator)
I28. We felt fully supported by management.	Motivation 1 (positive indicator)
I32. I knew why we were attempting this improvement initiative.	Motivation 1 (positive indicator)
I33. How many improvement initiatives had you participated in prior to this one?	'Previous Experience'
I34. Have you ever studied quality improvement methods and techniques?	'Previous Experience'
I38. Our initiative was successful in meeting its stated aim.	'Perceived Success'
I39. Our initiative achieved positive outcomes for patients, staff or the health system.	'Perceived Success'
I41. Please assign a score between 0-100 for your assessment of the success of your initiative.	'Perceived Success'

4.7 Pre-testing of the Survey Instrument

The survey instrument was pre-tested for timing, question wording, order, interpretation and basic content validity with six senior quality practitioner colleagues within the CDHB. Some minor clarifications of language resulted, such as the use of the term plan-do-study-act (the common usage in healthcare) rather than the plan-do-check-act more common in Lean literature. At the time of survey pre-testing it was also decided to use a five point Likert scale as opposed to a seven point Likert scale because some concern was raised over respondents' ability to reliably distinguish the intermediate levels of disagreement/agreement a 7 point scale would introduce. This decision was based on previous survey experience within the CDHB and the fact that the intended survey respondents were not specialists in QI language or practice. The subsequent pilot phase, with n=37 respondents, revealed that the 5 point scale did not inhibit a good separation of responses as there were a reasonable number of extremes (especially strongly agreeing responses) alongside agreeing (coded 4), indifferent (coded 3) and disagreeing (coded 2) responses. In addition, it has been shown by researchers that in practice, a seven point Likert scale offers no real benefit over a five point Likert scale, although in theory, a seven point Likert scale is more reliable than a five point Likert due to greater theoretical separation of the responses (Marsden & Wright, 2010).

4.8 Survey Delivery

A personal introduction from the respective programme leaders was included within the invitation to participate in the survey. All members of the of target survey populations were invited to respond by email and the survey was administered via the internet using Google Forms. A reminder email was sent 1 week before the nominal response window for each cohort closed.

4.9 Conclusion

Having adopted a positivist research approach, a survey was determined as the most practical method to collect data. Measurement items for each latent variable in the causal model were developed, using existing survey instruments where these could be identified. In the circumstances where no appropriate extant surveys instruments were identified, the survey items were developed from relevant and current theory associated with each construct. PLSPM was selected as the appropriate method for multiple regression data analysis. The next chapter outlines the results of the survey and data analysis results.

Chapter Five

RESULTS

5.1 Introduction

This chapter details the data collection and full analysis results. A summary of the data analysis steps is first presented in section 5.3 before each of the important test stages is detailed. Preliminary tests, PCA tests for unidimensionality, and initial reliability and validity testing are covered. Section 5.6 introduces the PLSPM results, including the structural model and a second check of measurement reliability and construct validity as determined by the modelling algorithm. The chapter concludes with a summary of the hypothesis test results.

5.2 Survey Response Rates

Even with an introduction from the respective programme leaders, the response rate for the survey was quite low, with an overall response rate of 15.4% (Table 5.1). However the responses that were received were in the majority of cases uniformly completed with minimal data gaps, providing some preliminary confidence that respondents had been diligent in completing the survey.

Table 5.1: Survey Response Rates

Sampling Cohorts	Survey Invites Sent	Responses	Response Rate
Group 1: Collabor8/Independent projects	266	45	16.91%
Group 2: Releasing Time 2 Care_OPH	165	19	11.51%
Group 3: Releasing Time 2 Care_CHCH	250	41	16.4%
Total	**681**	**105**	**15.4%**

Despite the low survey response rate, sufficient responses (105) were obtained to allow subsequent regression analysis; power analysis based on the methodology prescribed by Cohen (1992) indicated that the minimum sample size required for the study to achieve a 80% power for a medium effect size (medium R^2) is 97 cases. The down side of a smaller sample size such as 105 is that if there are weaker relationships in the data (structural causal paths), these will likely appear as non-significant (p >

0.05). This impact on weaker effects appeared to be the case with the moderator variables; the point is explored further in the limitations and methodological learnings section.

Group 1 was made up of 'graduates' of the Collabor8 programme (the short two-day course introducing Lean concepts and tools), as well as a small number of independent projects. Groups 2 and 3 were two separate iterations of the "Releasing Time to Care" programme — a more formally structured programme based on the NHS "Productive Ward" model. A statistically significant difference in group means for four of the seven constructs was noted (Table 5.2)[4]. Groups 2 and 3 (the Releasing Time to Care cohorts) had lower mean scores for the cause variables and higher mean levels for negative motivation. The p values <0.05 for the F tests (table 5.2) indicate that any assumption that the three groups have the same mean and distribution should be rejected i.e. the groups are dissimilar. This finding is not surprising when the makeup of the groups is considered. The interpretation is that Group 2, and to a lesser extent Group 3, had a higher number of disengaged or negatively motivated respondents. Given that these groups were the Releasing Time to Care cohorts, the question of instigation and sense of ownership of the improvement initiative affecting motivation is raised (Kaplan et al, 2013). The finding may well be of interest to the CDHB for evaluation of the programmes and the stratification may also offer clues for future research examining the source influences on the constructs (e.g. sense of ownership of the improvement initiative).

Table 5.2: Comparison of Group Means for each Latent Variable

	Lean Actions	**Team Work**	**Respect for People**	**Motivation _Pos**	**Motivation _Neg**	**Previous Experience**	**Perceived Success**
GR1 Mean	4.16	4.04	4.02	4.28	2.38	2.06	4.16
GR2 Mean	4.02	3.65	3.72	3.84	3.03	2.50	3.47
GR3 Mean	3.98	3.80	3.87	4.01	2.63	2.12	3.64
F test	$F = 1.29$ $p = 0.279$	$F = 4.34$ $p = 0.013$	$F = 1.50$ $p = 0.228$	$F = 3.98$ $p = 0.022$	$F = 4.96$ $p = 0.009$	$F = 1.50$ $p = 0.228$	$F = 8.20$ $p < 0.001$

[4] Ideally this information would have been presented graphically but the researcher experienced difficulty achieving a professional graphical output for this test in the R analysis software.

5.3 Data Analysis

5.3.1 Summary of Data Analysis Activity

After preliminary data checks and coding of the responses, PCA was conducted to assess the successful operationalisation of the hypothesised constructs. The PLSPM model was then built for the data analysis. Table 5.3 lists the sequence of data analysis activity carried out after the survey was completed. The major steps are then detailed.

Table 5.3: Summary of Data Preparation and Analysis Steps

Step	Data Preparation and Analysis Activity
1.	Survey results downloaded
2.	Survey iterations combined
3.	Transposition checks conducted on the combined data from 3 survey cohorts
4.	Survey results coded 1-5 (the reverse scored items were reverse coded at this step)
5.	Data frame created for analysis using the R software programme
6.	Initial data exploration
7.	Basic reliability and validity checks (some survey items removed at this step)
8.	Comparison of sub group means
9.	Principal Components Analysis; list-wise deletion method for missing data (some survey items removed at this step)
10.	Imputation of missing data using multivariate imputation. 3% of distinct responses imputed
11.	Principal Components Analysis using full imputed data set (further items removed at this step)
12.	Identification of principal components and measurement indicators to be retained
13.	Structural models programmed for partial least squares path modelling
14.	Measurement models programmed for partial least squares path modelling
15.	Path modelling conducted on regression models
16.	Repeated path modelling including bootstrapping
17.	The 'Negative Motivation' items, originally reverse coded, were restored to their original values
18.	Final iteration of path modelling including bootstrapping

5.3.2 Preliminary Data Analysis

Survey responses were combined into a master file and coded on a 5 point Likert scale. Preliminary checks were made on the completeness, range and basic consistency of the responses. One response was excluded due to the initiative still being substantially in progress and not ready for assessment. Missing data was imputed using multivariate imputation and rounded to match the Likert responses. The total number of distinct data imputed was 3% (116 out of 3675 answers in the final data set).

Prior to hypothesis testing (partial least squares path modelling), survey responses were subjected to PCA without factor rotation to ensure that multiple factors would emerge (Eigenvalue > 1.0). According to Harman's single factor test (Harman, 1976), if only one factor (dimension) emerges instead, this is suggestive of common method bias, because this implies that all measures have received a similar patterns of quick reflex responses to (wrongly) suggest that every survey item is related to every other item in the dimensional space. The PCA indicated 6 factors (Eigenvalues being 9.29, 2.64, 1.81, 1.35, 1.20, 1.10) based on Kaiser's criterion of Eigenvalue > 1.0 (Kaiser, 1960), thus suggesting absence of common method bias. The largest component accounted for 33 % total variance, less than the 50% level indicative of common method bias (Harman, 1976).

Harman's single factor test is regarded by many statisticians as an outdated and less than ideal testing method (Podsakoff, MacKenzie, Lee, & Podsakoff, 2003); however more advanced tests for common method bias were not available to the researcher. It is noted here that the survey construction was also designed to minimise these 'automatic' type of responses; the survey as delivered was divided into different segments for each variable, requiring fresh attention and a mental 'reset' from respondents.

5.3.3 Content Validity

Content validity was not specifically tested beyond the pre-testing of the survey with CDHB QI practitioners. A minimum level of content validity is supported by the generation of the operationalisation of the variables from the literature. For the key variables of 'Lean Actions', 'Teamwork', and 'Respect for People', additional confidence is provided by the alignment of the chosen measurement items with the QI practices index developed by Lemieux Charles et al., (2002), and the teamwork quality index developed by Hoegl and Gemuenden (2001).

5.3.4 Construct Validity

Construct validity refers to the validity of the operationalisation of the constructs (Nunnally & Bernstein, 1994). For this study, the survey instrument was used as the medium to operationalise the constructs and collect the data (each construct contained multiple indicators/survey questions). Convergent and discriminant validity were examined first at the PCA stage, and then again as part of the PLSPM analysis.

5.3.5 Bivariate Correlation Analysis

A preliminary test of discriminant validity was carried out using bivariate correlation analysis of standardised indicator loadings. The correlation matrix of indicator loadings and latent variables is attached as Appendix 2. In this matrix, each item should correlate strongly and positively to the related items within the construct (represented as coloured blocks), and more strongly to the assigned construct than any others. Three items from the correlation analysis (Appendix 2) require comment. Item 18 “We used facts and data to understand performance” appears to correlate strongly to ‘Lean Actions’ as well as ‘Teamwork’. The decision to include this item under ‘Teamwork’, as opposed to ‘Lean Actions’, was intentional in the scale development, on the grounds that whilst it certainly falls within the Lean domain, it is by no means unique to Lean and is a strong and established element of healthcare QI fully independent of Lean.

Item 9 “We maintained a focused effort” positions strongly within ‘Teamwork’ and also ‘Respect for People’, which is the related construct within the 2nd order construct of human resources capability. Item 32 (“I knew why we were attempting this improvement initiative”), and to a lesser extent the other indicators for the ‘Motivation’ construct, also correlate with the ‘Teamwork’ construct. These correlations may be illustrating the difficulties noted by both Lemieux Charles et al., (2002), and Hovlid and Bukve (2014), concerning the difficulty of accurately distinguishing an action from its context. An alternative interpretation may be that, given that we are breaking down *closely related* constructs, some degree of correlation between the defined variables of interest is expected. In this research context, the initial correlation analysis was deemed acceptable to progress to further analysis via PCA and PLSPM.

5.3.6 Reliability

Reliability relates to connectedness of the measures (intercorrelation) of a measurement scale (Cronbach, 1951). Reliability testing was carried out alongside PCA. Tests were conducted using R Studio software (R Core Team, 2014). The criterion for acceptance of the measurement scale items reliability was set as a Cronbach's alpha of 0.8. Additional reliability measures were also conducted as part of PLSPM analysis, but preliminary checking of the reliability was deemed to be a sensible pre-test of the PLSPM measurement model.

Table 5.4: Scale Reliability Using Cronbach's Alpha

Measurement Scale	**Std Alpha**
'Perceived Success'	0.87
'Lean Actions'	0.80
'Teamwork'	0.82
'Respect for People'	0.84
'Motivation'	0.77
'Previous Experience'	n/a

Analysis of the results for the 'Motivation' scale (Table 5.4) revealed several poorly worded and unreliable items. There were some compound questions that were inconsistently interpreted by respondents and also some items that simply did not fit the scale. Results for the modified scale were deemed acceptable at alpha = 0.77. The items for 'Previous Experience' did not conform to a Likert scale, nor to a consistent content domain (these items captured initiatives/projects, training courses and work environments). This scale was deemed unsuitable and not carried forward to the PLSPM analysis.

5.4 Principal Components Analysis Results on Unidimensionality

Prior to the hypothesis testing (partial least squares path modelling), survey responses were subjected to PCA using R Studio software (R Core Team, 2014). The purposes of PCA are to exclude any poorly loading measures as well as to ensure that the scales of the constructs (once any unreliable items are excluded) are unidimensional. Emergence of more than one dimension (factor/component within a PCA framework) for a given construct (latent variable) in the model raises questions over the validity of the operationalisation.

The preliminary criterion for acceptance of the measurement scale items was set as a Cronbach's alpha of 0.8. For the PCA, the acceptance criteria were an Eigenvalue for the principal component greater than 1 (Kaiser criterion), and a loading of the measurement item to the component of at least 0.7, following varimax rotation. When these criteria were not met the components were reviewed and unreliable/offending items were removed from the scales. In total, 14 items were removed from the measurement scales during reliability and PCA analysis. To an extent, this highlights the generic difficulty in constructing new measurement scales. It perhaps should also be noted here for completeness that the structural modelling treated the measurement items as reflective of the latent variables being measured (as opposed to formative). The objective when removing items was to optimise scale reliability and indicator loading, within the boundaries of the constructs as defined from the literature. Where a measurement item negatively impacted reliability/loading, but was not essential to the operationalisation, it was considered acceptable to remove the item. Full details of the final PCA iteration, including scree plots, un-rotated and rotated factor loadings are contained in Appendix 3.

Only two PCA results required subjective interpretation against the acceptance criteria. 'Lean Actions' appeared to have two principal components when judged solely on the Kaiser criterion. However no strong explanation for the groupings was discernible. Considering this lack of a valid reason to distinguish the components, and the significant Eigenvalue of component 1 relative to the other components, it was decided to continue to treat the 'Lean Actions' variable as a single construct (thus the construct was treated as unidimensional). The 'Motivation' variable showed only one Eigenvalue greater than 1. However the loadings patterns indicated retaining a second component would be desirable as the five measures get allocated to the two components in a statistically and conceptually meaningful way. The distinction between the items was based on whether the item was viewed as an indicator of presumed positive motivation (component 1) or presumed negative motivation (component 2). This presumption of positive or negative motivating influence is open to challenge and is discussed further in chapter 6 following the PLSPM results.

5.5 Partial Least Squares Path Modelling (PLSPM) Results

PLSPM was carried out using SmartPLS 3 Software (Ringle, Wende, & Becker, 2015). The outer model (the measurement model) was constructed using the response data from the final set of retained survey items (27 retained items). The inner model (the structural model) reflected the six research hypotheses; the moderator (contextual factor) 'Previous Experience' was excluded due to lack of sufficient data to test. Unfortunately, this also removed the possibility of testing the fourth hypothesis (H4). Following the first iteration of PLSPM, the moderator positive motivators did not demonstrate statistical significance ($p > 0.05$) and so this was also removed from the structural model. The final version of the model was then re-run with bootstrapping (2000 sample iterations). Key results are detailed and discussed below, and the complete SmartPLS report is attached as Appendix 4. However, before examining the hypothesised theoretical relationships (i.e. examining the structural model), the validity of the operationalisations of the constructs were re-tested via construct reliability and validity.

5.5.1 Construct Reliability and Validity (PLSPM)

The reliability of the measurement scales is a necessary but not sufficient condition for construct validity (Cronbach & Meehl, 1955; Messick, 1995). Although Cronbach's alpha is regarded as the primary measure of reliability, it is treated as a conservative measure on reliability because this coefficient relies on the Tau equivalency assumption — the assumption that all measures do have equal influence in shaping up the meaning of the construct (Chin, 1998; Hair, Hult, Ringle, & Sarstedt, 2014). The Tau equivalency assumption conflicts with the PLSPM algorithm (in PLSPM, each measure is assigned a specific weight to calculate the factor score and this contradicts the assumption of equal weight), and users of PLSPM are encouraged to use the alternative reliability measure known as *composite reliability* (ρ) (Hair et al., 2014). The prescribed cut-off values for coefficient ρ (composite reliability coefficient) reaching sufficient reliability are the same as those prescribed for reliability coefficient α (Cronbach's alpha): > 0.9 is excellent; > 0.80 is good; > 0.70 is adequate (Nunnally, 1978). Another measure of reliability — and more so the convergent validity — is the amount of variance a construct extracts from its assigned measures on average; this is termed the Average Variance Extracted (AVE) (Chin, 1998). AVE can also be interpreted as the average amount of variance of the indicators that the construct is able to explain—mathematically, the average of squares of the indicator loadings (Chin, 1998). An AVE of 0.50 (50%) or above is deemed satisfactory, although lower values can be considered provided there is other evidence to retain the measures (Chin, 1998; Hair et al., 2014). Table 5.5 depicts the reliability measures.

The second-order construct ('Human Resource Capability') is not listed in Table 5.5 because this construct is operationalised by its first order constructs ('Respect for People' and 'Teamwork'). Conceptually, if the first order constructs are reliable, then by definition, their second order construct is reliable. The reliability and convergent validity of the second order construct 'Human Resource Capability' can also be established from the indicator loadings (Figure 5.1) for the second order construct: 0.927 for 'Teamwork' and 0.841 for 'Respect for People'; these are high loadings. The implied AVE value of the second order construct is 0.783 (the average of 0.927^2 and 0.841^2), which again is high.

Table 5.5: Comparison of Reliability Measures for the Retained Constructs

Construct	**Cronbach's alpha**	**Composite Reliability (ρ)**	**Average Variance Extracted (AVE)**
'Negative Motivation' (Motivation_Neg)	0.684	0.863	0.759
'Lean Actions'	0.796	0.851	0.450
'Perceived Success'	0.866	0.918	0.789
'Respect for People'	0.849	0.898	0.688
'Teamwork'	0.819	0.869	0.526

All constructs showed evidence of good composite reliability. Although the moderator (Motivation_Neg) construct returns a Cronbach's alpha value that is slightly less than 0.70, this was ignored because of the Tau equivalency issue mentioned above. 'Lean Actions' return AVE values below 0.50. However, 'Lean Actions' return high values for Cronbach's alpha and composite reliability. For these reasons it was concluded that the measurement scales of all the constructs showed satisfactory levels of reliability.

5.5.2 Convergent Validity

Convergent validity refers to measures assigned to a particular construct being strongly related (having a high shared variance/convergence) to that construct (Hair et al., 2014; Straub, Boudreau, & Gefen, 2004). High indicator loadings (> 0.70 for each indicator is desirable) and a high AVE of the construct (> 0.50 ideally) are used as indicators of convergent validity (Chin, 1998; Hair et al., 2014; Straub et al., 2004). As shown in Table 5.5, the indicator loadings returned satisfactorily high values. It would have been desirable to receive higher loadings for some indicators of 'Lean Actions', but eliminating indicators solely for the sake of increasing the indicator loading values (and thereby the AVE) is not recommended in PLSPM as this compromises the meaning of the construct (Chin, 1998; Hair et al., 2014).

Table 5.6: Indicator Loadings of the Measurement Model

Latent Variable	**Indicator**	**Indicator Loading**
'Lean Actions'	I1	0.687
	I2	0.657
	I3	0.613
	I4	0.606
	I5	0.676
	I6	0.750
'Respect for People'	I11	0.823
	I12	0.845
	I14	0.801
	I15	0.850
'Teamwork'	I16	0.728
	I17	0.644
	I18	0.751
	I19	0.780
	I20	0.726

Latent Variable	Indicator	Indicator Loading
	I9	0.717
'Perceived Success'	I38	0.906
	I39	0.904
	I41	0.854
'Positive Motivation'	I21	0.875
	I28	0.673
	I32	0.849
'Negative Motivation'	I22	0.852
	I27	0.890
For all indicators, $p < 0.001$		

The AVE values of the constructs (Table 5.5) also returned satisfactory values, although again, it would have been desirable to have returned a slightly higher AVE for 'Lean Actions'. Based on the indicator loadings and the AVE values it was concluded that convergent validity had been established. Having demonstrated scale reliability and convergent validity, the remaining aspect of construct validity to test was the discriminant validity (Hair et al., 2014).

5.5.3 Discriminant Validity

Discriminant validity refers to sufficient levels of discreteness of the measures — that is measures being assigned to a construct being more strongly correlated with that construct than with the other constructs (Straub et al., 2004). The most straightforward examination of discriminant validity is the correlation between a measure and its assigned construct (i.e. indicator loading) relative to the correlations between measures and remaining constructs (i.e. indicator cross-loadings) (Chin, 1998). However this approach is no longer recommended, because simulation studies show that it has the tendency to clear measures as being valid too casually (Hair et al., 2014). Therefore, instead of this test (the loading cross loading test), the Fornell-Larker criterion was used (Fornell & Larcker, 1981). The Fornell-Larker criterion for discriminant validity is used to show that a construct shares more variance with its assigned construct (as indicated by its AVE) than with the other constructs, as indicated by the squared correlations between constructs. When examined as a matrix, the squared correlation for the construct should exceed the squared correlations of the remaining constructs, suggesting that the measures belong to their assigned construct as opposed to other constructs (Fornell & Larcker, 1981). This is mathematically equivalent to treating a construct as being more correlated with its measures (indicated by the square root of AVE) than with other constructs (Hair et al., 2014). The correlations in Table 5.7 indicate that the measurement scales possess adequate levels of discriminant validity because the value of the diagonal element (the square root of the AVE of the corresponding construct) exceeds the corresponding off diagonal elements — for example, for the construct Motivation_Neg, **0.871** > 0.308, 0.568, 0.338, 0.530.

Table 5.7: Correlations Between Constructs in terms of the Square Roots of AVE

Construct	**‘Negative Motivation’**	**‘Lean Actions’**	**‘Perceived Success’**	**‘Respect for People’**	**‘Teamwork’**
‘Negative Motivation’ (Motivation_Neg)	**0.871**				
‘Lean Actions’	0.308	**0.671**			
‘Perceived Success’	0.568	0.631	**0.888**		
‘Respect for People’	0.338	0.279	0.373	**0.830**	
Teamwork	0.530	0.617	0.637	0.577	**0.726**

Having established scale reliability and construct validity, the estimated structural regression coefficients and other related parameters of the PLSPM results were then examined.

5.6 PLSPM Parameter Estimates

Path coefficients (including the statistical significance), and the R^2 of the endogenous (dependent) constructs of the final model are shown in Figure 5.1. Note that the effect of the moderator 'Positive Motivation' (Moderator 1) was not included in the final analysis because the corresponding iteration term was found to be non-significant (T = 0.799; p = 0.426). Unfortunately, this also removed the possibility of testing the fifth hypothesis (H5).

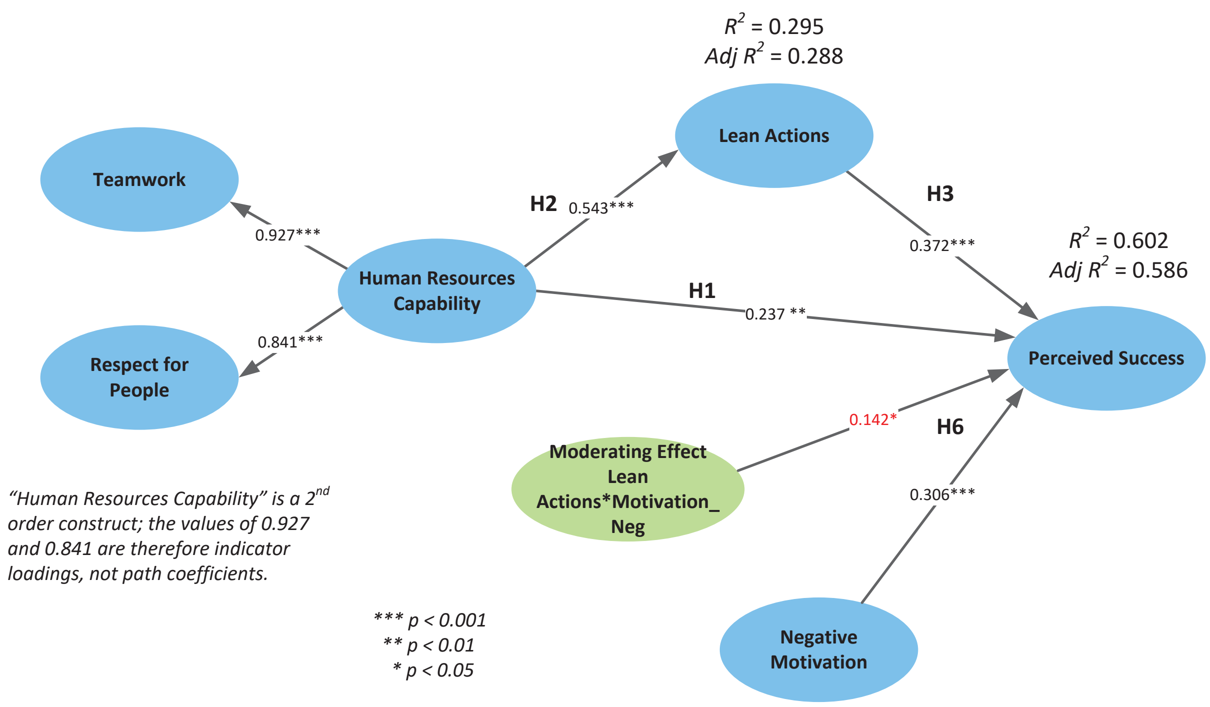


Figure 5.1: PLSPM structural model

The adjusted R^2 associated with the endogenous construct 'Perceived Success' returns a value of 0.58 (Figure 5.1), which is deemed quite acceptable in social research (Hair et al., 2014). However, the adjusted R^2 associated with the remaining endogenous construct 'Lean Actions' returns a much lesser value of 0.288, which is not a strong relationship. Nevertheless, all the path coefficients associated with this endogenous construct are significant, suggesting validity of the structural model. It is now possible to use the structural model obtained to examine which hypotheses were supported and which were not. The results are shown in Table 5.8.

Table 5.8: Hypotheses Test Results

Hypothesis	Supported?	Justification
H1: 'HR Capability' has a direct positive effect on 'Perceived Success'.	Yes	The corresponding path coefficient was found to be positive (0.237) and significant ($p < 0.001$).
H2: 'HR Capability' has a positive effect on 'Lean Actions'.	Yes	The corresponding path coefficient was found to be positive (0.543) and significant ($p < 0.001$).
H3: 'Lean Actions' have a positive effect on 'Perceived Success'.	Yes	The corresponding path coefficient was found to be positive (0.372) and significant ($p < 0.001$).
H4: 'Previous Experience' has a positive effect on 'Perceived Success'.	Cannot determine	There were not enough data to validly operationalise the moderator 'Previous Experience'
H5: 'Positive Motivation' has a positive moderating effect on 'Lean Actions' being able to improve 'Perceived Success'.	No	The corresponding interaction term was found to be non-significant ($T = 0.799$; $p = 0.426$) and was thus excluded from the final model.
H6: 'Negative Motivation' has a negative moderating effect on 'Lean Actions' being able to improve 'Perceived Success'.	No, but the possible ways negative motivation might motivate workers is discussed	The path coefficients of the predictors, including the corresponding interaction term were found to be significant ($p < 0.05$); However the interaction term coefficient has a positive coefficient, suggesting a positive moderation effect. This contradicts the hypothesis (the results imply that negative motivation may be beneficial.

5.7 Conclusion

Comprehensive tests were conducted on the survey data that was collected. PCA was used alongside PLSPM to examine the key reliability and validity tests prior to hypothesis testing. No contradictory or illogical findings emerged; however the convergent validity for the 'Lean Actions' construct fell marginally short of the nominally preferred level. When all the test results were considered as a whole this finding was deemed acceptable, given the acknowledged sample size limitations. The reliability and validity tests provided confidence in the integrity of the structural and measurement models used for PLSPM, therefore allowing the resulting parameter estimates to be used to test the hypotheses. The next chapter is a detailed discussion of the PLSPM results shown in this chapter.

Chapter Six

DISCUSSION

6.1 Introduction

In the theoretical model there were three types of constructs, all operationalised as directly unobservable variables (latent variables) using multiple indicators (measures). The first variable was the *causal variable*, namely, 'Human Resource Capability'; it is this 'Human Resource Capability' that causes successful outcomes. 'Lean Actions' also functioned as a causal variable, with the additional role of *mediator*; acting as the mechanism through which 'Human Resource Capability' caused successful outcomes. The second type of variable was the *effect variable* (the actual outcome), which is *Success*. This study used a proxy for Success — 'Perceived Success' — because it was not possible to directly capture the outcome of the QI initiates (the associates' perception on the outcomes was used). The third type of variable was the contextual variable, whose relative strength could amplify or attenuate the QI initiatives (i.e. the cause effect relationships). In terms of psychometric research, these variables are known as moderators. At the commencement of the research, three moderators were identified — 'Previous Experience', 'Positive Motivation' and 'Negative Motivation'. One moderator ('Previous Experience') had to be eliminated on practical grounds due to lack of data while the 'Positive Motivation' moderator was eliminated on the grounds of statistical non-significance (Table 5.8).

6.2 Direct and Indirect Effects

H1: 'HR Capability' has a direct positive effect on 'Perceived Success'

As shown in Table 5.8, the first hypothesis (H1) was supported. The path coefficient between 'HR Capability' → 'Perceived Success' indicates the direct effect, which is 0.237. The direct effect of 'HR Capability' on 'Perceived Success' is somewhat intuitive and certainly this finding is consistent with the earlier work of Lemieux-Charles et al. (2002); Hoegl and Gemuenden (2001) and Kaplan et al. (2012, 2013).

However, it is important to compare the above direct effect with the indirect effect through the mediating variable 'Lean Actions'. The indirect effect is caused through the following path: 'HR Capability' → 'Lean Actions'→ 'Perceived Success'. The indirect effect of 'HR Capability' on 'Perceived Success' can be estimated by multiplying the two path coefficients corresponding to the 'HR Capability' → 'Lean Actions' path and the ‘Lean Actions’ → ‘Perceived Success’ path. These two paths correspond to the second and third hypotheses.

H2: ‘HR Capability’ has a positive effect on ‘Lean Actions’

H2 has also been supported. The path coefficient representing ‘HR Capability’ → ‘Lean Actions’ is 0.543, which is a sizable effect. Such a sizable effect is required to support the proposal that quality improvement practices (in this case, Lean) mediate the causal path between ‘HR Capability’ and improvement outcomes, although the actual mediation effect (hence the indirect effect) depends on the remaining path (the path corresponding to the third hypothesis) also.

H3: ‘Lean Action’ has a positive effect on ‘Perceived Success’

The third hypothesis (H3) has also been supported. The path coefficient corresponding to ‘HR Capability’ → ‘Lean Actions’ is 0.372, which is however a moderate effect. Thus the indirect effect of ‘HR Capability’ on ‘Perceived Success’ through the Mediating Variable ‘Lean Actions’ is 0.543*0.372, which is 0.202. This can be compared with the direct effect of ‘HR Capability’ on ‘Perceived Success’, which is 0.237. As a result the total effect of ‘HR Capability’ on ‘Perceived Success’ becomes 0.439 (0.237 + 0.202). This finding of a higher direct effect of ‘HR Capability’ on ‘Perceived Success’ (albeit marginally) relative to the indirect effect due to ‘Lean Actions’, somewhat contradicts the findings of the empirical study on Toyota by Jayamaha et al. (2014). These researchers found that in Toyota, ‘Lean Actions’ almost fully mediate (≈100% mediation) the effect of ‘HR Capability’ on ‘Perceived Success’, meaning almost a zero direct effect of ‘HR Capability’ on ‘Perceived Success’, in the case of Toyota. The present study found that ‘Lean Actions’ only partially mediates (46% mediation, being 0.202/0.439) the relationship between ‘HR Capability’ (cause) and ‘Perceived Success’ (effect). Comparing Toyota and the CDHB as sociotechnical systems, there are distinctions that may explain this difference in mediating effect. Firstly, Lean is only one methodology in the pluralistic improvement landscape that exists in the secondary healthcare environment. Improvements may have been achieved (or perceived to have been achieved) by other actions e.g. a change in clinical practice. In addition, despite some familiarity and exposure to Lean concepts within the CDHB, this level of exposure is unlikely to approach the level of individual assimilation of these concepts that exists within Toyota. An action such as the introduction of a new checklist could be interpreted differently in each culture. It is easy to define such an example as a Lean Action (standard work), but it may not be recognised as such by an inexperienced Lean practitioner (the majority of the research participants). Potentially the relative inexperience of the survey population has underestimated the use of ‘Lean Actions’, and thus reduced the mediating effect as measured.

6.3 Moderating Effects

H4: ‘Previous Experience’ has a positive effect on ‘Perceived Success’

As previously noted, the hypothesis that ‘Previous Experience’ has a positive effect on ‘Perceived Success’ was not tested, due to inadequate data. The moderating variable ‘Previous Experience’ proved difficult to measure. The critical obstacle encountered was determining the relevant domain of experience to attempt to measure. As an abstract concept, ‘Previous Experience’ could be legitimately viewed for different purposes as encompassing any mix of tenure, training, qualifications, life skills, depth of knowledge, or breadth of knowledge, which may explain the inconsistency of responses. Failing to account for any of these particular domains adequately risks content validity and mixing the domains poorly compromises construct validity (DeVellis, 2012).

H5: ‘Positive Motivation’ has a positive moderating effect on ‘Lean Actions’ being able to improve ‘Perceived Success’

The Hypothesis that positive motivators have a positive effect on ‘Perceived Success’ was not able to be supported, because the results did not reach significance and the null hypothesis could not be rejected. There are a number of potential interpretations of this finding. Firstly, the null hypothesis may be correct — i.e. H5 is false. A second possibility is the sample size at n=105 is small and may be insufficient to highlight weak effects, particularly when there is some collinearity between variables, i.e. the *amount* of data for this variable was not suited for regression analysis as a moderator (Cohen, 1992). A third possibility is that the measurement scale may not have adequately reflected respondents’ positive motivation to invest in the QI activity i.e. despite the efforts to build the construct from applicable theory, it still failed *content* validity. A final consideration is the influence of positive respondent bias in the survey sample, given that respondents self-selected and were presumably intrinsically motivated to respond. Overall, the survey responses were somewhat positively biased in the sense of strong levels for the predictor variables (i.e. associates self-reporting strong

‘HR Capability’ and ‘Lean Actions’), and high scores for the outcome (‘Perceived Success’) variable. In simple terms, the majority of respondents considered their initiatives to be successful. This bias can be seen in Figure 6.1, which is a simple distribution of the ‘Perceived Success’ score (Item 41):

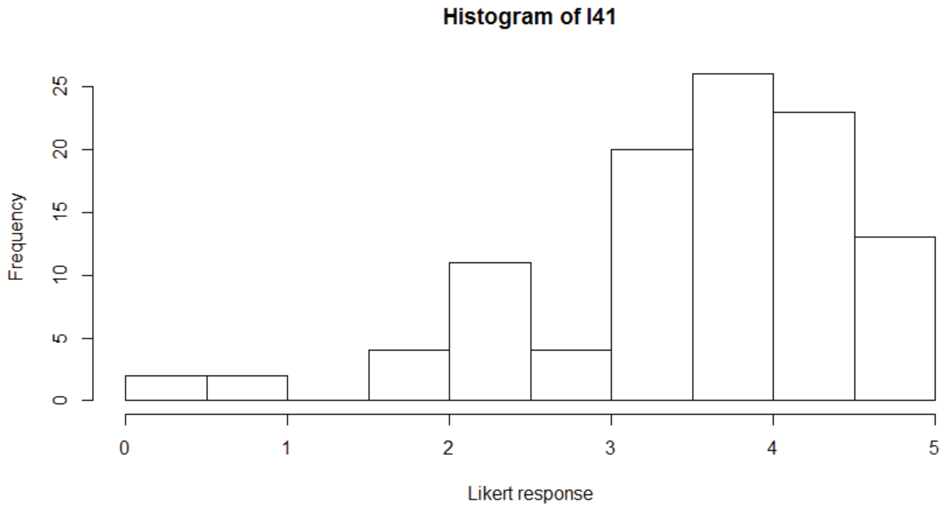


Figure 6.1: Distribution of Item 41 (‘Perceived Success’) Score

Potentially, a high number of the survey respondents were well motivated generally and sufficiently capable to achieve good results; any smaller moderating effects from additional motivation that could be ascribed to the ‘meaningful rationale of goals’ theory (Gagné & Deci, 2005) may have been too small to distinguish in this group.

H6: ‘Negative Motivation’ has a negative moderating effect on ‘Lean Actions’ being able to improve “Perceived Success”

H6 was not supported. The results showed a slight, but statistically significant, result of 0.142, which does not suggest a negative moderating effect. This finding was unexpected because it was the opposite of the hypothesis. The implication is that the presence of negative motivators (slightly) improves outcomes. Figure 6.2 shows the moderator slope analysis:

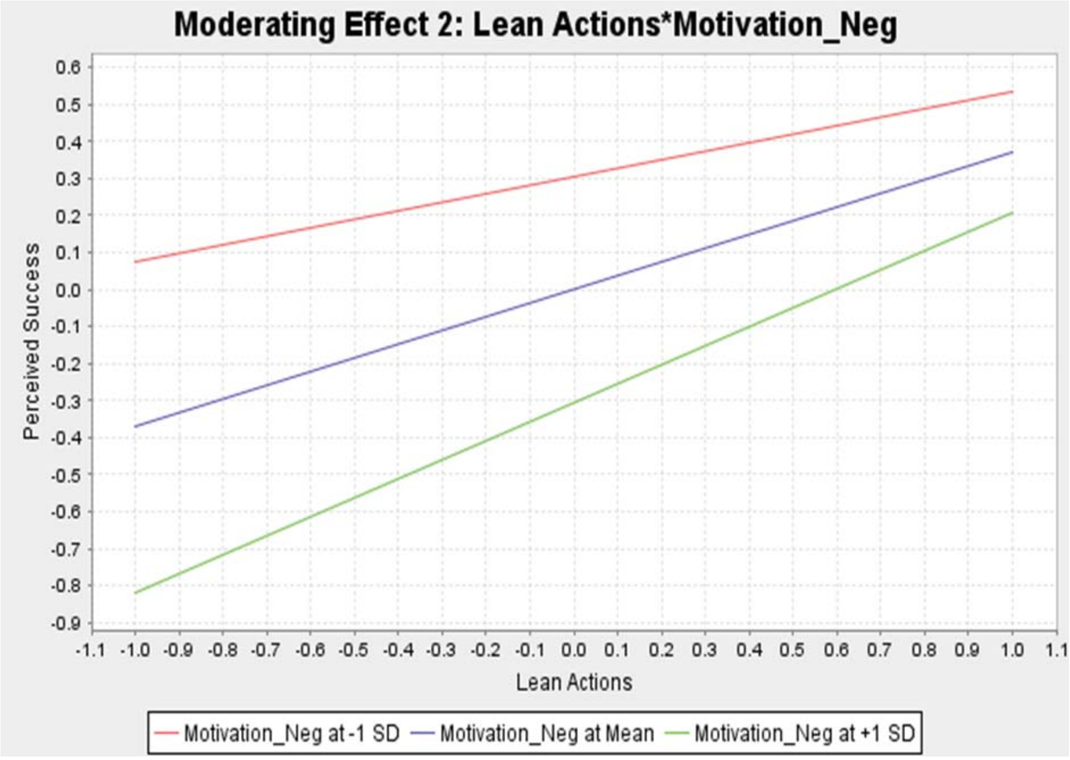


Figure 6.2: Moderating Effect of 'Negative Motivation'

In Figure 6.2, the green line represents negative motivation being present (the negative motivation indicators set to the highest level = *negative motivation is present*; refer to footnote[5]). Although only a slight effect, the green slope is discernibly steeper than the red line (the negative motivation indicators set to the lowest level = *negative motivation is not present)*. The steeper slope of the green line suggests a stronger moderating effect on 'Lean Actions' when having some negative motivators present (e.g. setting goals for the associates, which in their opinion are unrealistic). Thus, as opposed to no such negative motivators being present, the presence of the negative motivators causes associates to produce slightly *greater* success for given levels of Lean Action.

Given that the survey pretesting, PCA and PLSPM tests all supported measurement reliability and validity for the 'Negative Motivation' construct, the researcher takes the position that the result could equally be correct and therefore the hypothesis requires refinement for further testing with a new dataset, ideally under a different Lean

[5] In the early development of the PLSPM model, this moderator was originally operationalised as "absence of negative motivators". The double negative language proved cumbersome and confusing in relation to the hypothesis, and so the moderator was eventually relabelled more simply as 'Negative Motivation'. For the final iteration of the PLSPM algorithm results (as shown in figure 5.1), the original indicator scoring for this variable was reinstated, to allow the hypothesis to be tested as stated. This process is detailed further in the methodological learnings section 7.6.

environment in healthcare. The proposition that for at least some of the respondents, negative motivators in some way helped them succeed has support from theory. For example, the central importance of challenge and difficult goals in Lean is noted by Womack and Jones (2003) and the requirement for 'difficult but fair' goals is a recognised part of goal setting theory (Gagné & Deci, 2005). It is not unreasonable to expect that some teams or individuals respond well to challenging circumstances, increasing motivation to overcome perceived obstacles to success. H6 as stated may still correctly apply to some teams or individuals, but it may be failing to account for this 'overcoming challenges' effect. It would be beneficial to conduct case studies (qualitative data collection) to further understand the associates' attitude and perception towards QI work goals in Lean healthcare, to refine the hypotheses on motivation (both positive and negative) before these are tested with quantitative data.

6.4 Conclusion

The PLSPM parameter estimates convincingly support hypotheses 1, 2 and 3, all of which relate to the direct and indirect paths of the cause-effect relationship. The magnitude of the mediating effect from 'Lean Actions' was noticeably less than that found by Jayamaha et al. (2014) in their study of the Toyota sociotechnical system. The difference in the relative level of internalisation of lean concepts between participants in the two studies is suggested as the most likely explanation for this finding. Hypothesis 4 was not tested due to inadequate data. Hypothesis 5 was not supported due to the results not reaching significance. Results for hypothesis 6 did reach significance, and contradicted the hypothesis. This unexpected finding suggests additional research is required to fully understand how to test the 'Motivation' contextual factor.

The next chapter concludes the main body of the thesis by taking a retrospective look at the findings of the study in light of the original objectives.

Chapter Seven

CONCLUSION

7.1 Introduction

In this chapter the outcomes of the research are compared back to the original research objectives. Section 7.5 then revisits the limitations of the research in more detail, and section 7.6 sets out the methodological learnings the researcher has taken from the study. The chapter concludes by bringing the findings, limitations and learnings together into suggestions for the next stage of the research.

7.2 Reviewing the Outcomes Achieved Against the First Research Objective

The first research objective was "*To identify and measure the prevalence of contextual factors that affect quality improvement initiatives in New Zealand public hospitals*".

The literature review revealed that the most common and/or critical contextual factors for QI in healthcare and other industry sectors have been identified (or at least proposed). However the descriptors and operational definitions are not consistent and a key theme of the literature is to improve this situation in current and future research (Damschroder et al., 2009, Øvretveit 2010). Within healthcare, multiple contextual models have been developed and are in use by researchers (Andersen et al., 2014, Kaplan et al., 2013). However, the present study has highlighted a key limitation of these models in that they do not adequately specify the relationships between the contextual factors, and this is a precondition to meaningful measurement. Therefore a causal model was developed to allow measurement of selected factors. A selection process from within the full list of possible contextual factors was required for two reasons. Firstly, the preliminary work of Kaplan et al. (2013) on MUSIQ strongly suggested the importance of factors operating at the microsystem level, therefore attention was focused on the microsystem. Secondly, the existing contextual factor models, including the one selected as the basis for identifying contextual factors (MUSIQ), juxtapose a wide range of factors that in reality do not interact at the same level of activity. Higher level factors, such as organisational culture and leadership, are combined with much more narrowly prescribed factors such as individual QI skill. Supported by the literature, the present study highlighted this mixing of levels within the existing contextual factor models.

The importance of leadership, organisational culture and other elements that might be termed "Business Excellence" fundamentals is well known to researchers and practitioners. The roles these vital organisational attributes play were not ignored in this study; however they were positioned as background exogenous variables which undoubtedly influence the microsystem hypothetical model, shaping and influencing 'HR Capability'. This study joins the previously noted studies in confirming and reinforcing the central importance of 'HR Capability' on QI; the important organisational influences affecting 'HR Capability' therefore remain critical to explore further. Given the positivistic research paradigm underpinning the study, the researcher maintains that the first research objective has been successfully achieved by way of providing a scientific basis for the selection of contextual factors. A basis to accept or refute (through hypothesis testing) the explanations for inclusion of each contextual factor has been developed.

7.3 Reviewing the Outcomes Achieved Against the Second Research Objective

The second research objective was "*To explain the key empirical relationships between quality and process improvement interventions, outcomes and the contextual factors, from a theoretical and practical standpoint*".

The final results for this study illustrate the full range of possible outcomes from hypothesis testing: three hypotheses were supported, one hypothesis was unable to be tested due to the data collected, one hypothesis was non-significant and one hypothesis was inconclusive. With reference to MUSIQ, the results fully support the causality as proposed in the MUSIQ model — i.e. *QI Capability* and *Team Norms* (in the language of MUSIQ), positively influence system and process change (Kaplan et al., 2012, 2013). This finding is consistent with the findings of other researchers including Lemieux-Charles et al., (2002); Hoegl and Gemuenden, (2001); and Jayamaha et al., (2014), where the causal link between teamwork, QI practices and improvement outcomes was established. An important task for healthcare QI researchers remains the consolidation of the terminology and structure of their conceptual models; this will allow all parties to be confident they are discussing and measuring the same thing (Damschroder et al., 2009; Portela et al., 2015).

New knowledge was added to one prominent model by testing MUSIQ in a Lean environment to explain the effect of contextual factors on outcomes. This was achieved by treating 'Lean Actions' as the mediating variable in the model, i.e. the mechanism of process change. Although Lean is a complex domain to reduce to just two latent variables — one representing 'Human Resource Capability' and the other representing the actions of this intangible resource (Lean Action) — it was able to be measured successfully. The findings on teamwork and capability have a plausible basis in Lean theory, in the 'Respect for People' pillar of the Toyota Way, the intangible resources and soft skills of an organisation (Jayamaha et al., 2014). The research also demonstrated that abstract contextual factors that are not directly observable can be measured in healthcare microsystem settings, although admittedly not without some difficulties as noted in the previous section.

Referring back to the literature and the recognition of the difficulties involved in successfully implementing Lean in healthcare QI (Andersen et al., 2014; D'Andreamatteo et al., 2015; Moraros et al. 2016), this fundamental relationship between team capability, team interactions and improvement outcomes cannot be overlooked. The recognition of the relationship is not limited to healthcare, these are precisely the so called 'soft' Lean practices now being examined closely in Lean implementations across other industry sectors (Bortolotti, et al., 2015). The strong conclusion is that it is essential for organisations and QI practitioners to genuinely recognise the importance of supporting human resources capability alongside any specific Lean tools when implementing Lean improvement initiatives. Thus, the researcher maintains that the second research objective has also been successfully achieved.

7.4 Reviewing the Outcomes Achieved Against the General Research Objective

The general research objective was "*To propose guidelines for quality improvement in the secondary healthcare sector, based on the empirical study findings*".

Statisticians commonly caution that "all models are wrong" at some level of analysis (Crawley, 2015, p.193), and the limitations of the model presented here have been fully acknowledged. Nevertheless, the clear outcomes for the first three of the hypotheses, fully consistent with theory and previous empirical findings, are very encouraging. The findings also suggest that a convergence of multiple context-sensitive models for healthcare quality improvement is achievable; i.e. separately developed models that share underlying theory can be brought together in a way that retains construct validity and allows for wider quantitative analysis. However the contextual models must first be explained in terms of interacting variables as is standard for social science research. A second key learning is that the quantitative analysis must be at sufficient scale to measure the weak effects of contextual factors functioning as moderators of a cause-effect relationship. This study highlighted the dominant effect of the 'HR Capability' construct (which combines 3 contextual factors from the MUSIQ model: *Team Norms, Skill and Decision Making*). In one practical sense, the finding that the moderating effects appear to be much weaker than the dominant cause-effect relationship has value in its own right. Exploring the 'HR Capability' contextual factors in more depth is potentially of more interest to QI practitioners than spending time on the weaker moderating factors.

All parties interested in healthcare QI research share a desire to better understand how their improvement efforts can produce the best results and how to reliably assign scarce resources. A practical predictive tool, in the form of a concise, validated survey instrument capable of assessing QI activity in progress, although not delivered directly from this single study, remains a realistic goal. This study has supported previous findings highlighting the criticality of effective teamwork, communication and interaction amongst QI participants, so there is a validated starting point for QI practitioners to focus their support efforts on. Earlier recognition of these contextual challenges, along with isolating the subsequent key interventions indicated to have the most influence, will help to keep initiatives on track for success. The present study treated higher level contextual factors such as organisational leadership as necessary fundamentals in line with the established and well researched principles of Business Excellence Frameworks. The problem for operational staff and QI practitioners is that they have very limited ability to influence these critical success factors. They can however exert considerable influence over their teamwork, respect for people and QI capability, so the potential value of further research into how ‘HR Capability’ drives successful outcomes remains very high. Finally, it is important to not let the detailed, tightly prescribed focus required for this study obscure the true purpose of the research — ensuring that patients will benefit now and in the future from a health system more capable of responding to their needs quickly via effective QI practices.

7.5 Limitations of the Study Revisited

Several limitations of this study are acknowledged. The sample was a nonprobability sample —participants of the chosen improvement programmes were invited to respond and self-selected. Despite efforts to encourage responses, the response rate was low and respondent bias was introduced via self-selction.

The self-reporting of initiatives without any independent data source also introduced the risk of common method bias. Although the PCA suggested that this bias was avoided, other implications of the lack of an independent data source should still be considered if the research is repeated in another setting. As one example, Lemieux-Charles et al., (2002) showed wide divergence between QI participants and management when reporting on improvement outcomes. However for the purposes of the present study, respondents’ perceptions of their activity were the item of interest.

Some potential implications of the restriction to the microsystem layer of the MUSIQ model need to be noted. For almost all of the survey respondents, the initiation and the control of the improvement activities remained within their very small working team. If the surveyed initiatives had involved larger, specially assembled project teams, requiring greater cross-organisational cooperation, the results might not have been the same. We can assume a wider range of perceptions and behaviours within these larger ‘teams’ and more contextual factors influencing the structural model.

In the MUSIQ model a distinction is made between contextual factors operating at the QI team level (e.g. a project team) and the microsystem (staff carrying out the process changes). This distinction between ‘project’ and ‘everyday’ work was not well demarcated by survey respondents and could not be reliably distinguished in the results.

This may be a potential limitation of the MUSIQ model in practice, but it is also possible that this is simply a result of the focus on microsystems where staff are working on their own systems and processes and effectively functioning at both levels.

7.6 Methodological Learnings

7.6.1 Survey Learnings

Although the survey was able to be easily completed within the planned 20 minutes, this may still have been seen as too long or too much effort by many respondents. The response rate strongly suggests alternative strategies would be advisable for future research involving these target groups. Some feedback received from respondents suggested a fundamental issue with the decision to use an online survey. Access to computers during working hours for many of these respondents, especially the nursing and allied health staff, was often limited and any online activity consequently highly prioritised during working hours.

The time window for applicable QI initiatives allowed as part of the inclusion criteria (up to two years) may also have played a role. For some respondents, their willingness to invest time and effort in thinking about a small scale improvement initiative undertaken two years ago may not have seemed worth the effort (beyond the request for support and appeal to organisational learning from the programme leaders, participation was not explicitly requested from senior management and no other incentives were offered). Unfortunate timing may also have contributed for the lowest responding cohort, who were in the process of preparing to move hospitals. This group also had the most limited access to computers during their work shifts. Finally, wide variation in staff engagement levels with the respective programmes may have contributed, but examining those constraints further was outside the scope of the research.

The survey exemplified a number of the known challenges to effective survey research, especially if using a newly developed survey instrument. Although this process can be considered pretesting and validation for a new survey instrument, it is probably a small research project in its own right. An early decision was made to include some statements within the 'Teamwork' and 'Motivation' constructs despite not having examples from previous instruments or a sufficiently strong theoretical foundation for them. The intent was to 'uncover' the content domain or additional constructs via respondents' answers. In hindsight this was not an effective tactic, as a number of these responses did not align strongly enough with the latent variable being measured. The result was a survey longer than it needed to be; with responses that weren't able to be included in the quantitative analysis (i.e. data had been collected at some effort that could not be used). In the end, 27 out of the intended 41 statements were used in the quantitative analysis.

Overall, the individual statements appear to have been worded well, although a few compound statements were unintentionally retained in the survey and these proved problematic to respondents. An example is the statement for Item 31: "I believed in this initiative, despite not knowing how to achieve successful outcomes". The responses to this statement were unreliable and this item failed the initial reliability testing. Novice researchers are forewarned of these problems in standard texts on survey construction (e.g. Bryman, 2012; DeVillis, 2012); the present study provided a 'textbook' illustration of the warnings.

7.6.2 'Motivation' Construct

Of all the constructs considered in the hypothesised model, the 'Motivation' construct proved to be the most complex. This was partly due to the wide range of possible theory to consider as the foundation, and also due to the potentially different interaction mechanisms. Despite the significant body of research on motivation in work and organisational settings in general (well summarised in Gagné & Deci, 2005), strongly supported theory as applied to motivation for (and subsequent impact on), quality improvement interventions is limited. Given that the "how" of the role Motivation plays in QI is not clear, this suggests another research strategy may be required to complement the quantitative approach taken in this study (Yin, 2013). Qualitative research, via case studies with healthcare QI practitioners, may provide an improved understanding of QI practitioners' motivation and how this affects their work, and a stronger foundation for the hypothesis formulation.

As previously noted, the sample size had adequate statistical power for moderate to strong effects but low statistical power for weak effects (Cohen, 1992). The moderators are not hypothesised to have strong effects in relation to the other variables, and so therefore a larger sample size is required to test these adequately. A sample size of at least 700 will be required (Cohen, 1992, p. 158.).

For completeness, the unusual labelling of the motivation constructs is fully explained here. Initially, motivation was conceived as one single construct that would combine the positive and negative scores into a single composite score of “relative motivation”. To allow for this aggregation, the items measuring negative motivation had to be reverse coded (so that an item indicating strongly positive motivation, coded 5, could be combined with another item indicating strongly negative motivation, coded 1, for a meaningful overall score). The PCA analysis supported the positive and negative directions of the motivators, but it was more practical to separate the construct into two for the PLSPM modelling of moderating variables. Negative motivation, reverse coded, thus became “absence of negative motivation indicators”, with results coded as 5 representing very little negative motivation present and results coded 1 representing a lot of negative motivation present. Unfortunately, to test Hypothesis 6 (‘*Negative Motivation’ has a negative moderating effect on ‘Lean Actions’ being able to improve ‘Perceived Success’*) correctly, an interaction effect of ‘Lean Actions’ * ‘Negative Motivation’ working in the same direction is required to reveal the correct amplitude of any interaction effect and the final pos/neg direction of this effect. For the first (preliminary) iteration of PLSPM, the direction of the estimated path coefficient for this interaction effect was incorrect in terms of testing the hypothesis. A final iteration of PLSPM was therefore run with the indicator scores for the ‘Negative Motivation’ measurement items reverted to their original (non-reversed) values. This restored the direction of the estimated path coefficient from - 0.142 to + 0.142 (reflecting ‘Lean Actions’* ‘Negative Motivation’ (correct) rather than the ‘Lean Actions’* Absence of ‘Negative Motivation’ (incorrect). All other estimated parameters remained identical to the original PLSPM iteration.

Although not ideal, this process was carefully managed at each stage and has been tracked through the entire process, from survey data collection through to PLSPM testing. It was another valuable methodological lesson that also raised a potential question for future similar research on motivation *direction*; whether it is preferable to treat positive and negative motivation as separate constructs or combine them into a single, averaged or aggregated construct at the measurement scale stage (i.e. measuring in both negative and positive directions, versus measuring from zero to positive only, with zero representing the strongest negative score).

7.7 Recommendations for Further Research

This final section considers knowledge gaps and research aims for future research on this topic. Treating the present study as an initial proof of concept, there are many opportunities to develop the research further, in particular, examining the interaction of motivation on QI capability and teamwork in more detail, where the results were inconclusive. As previously noted, this will require a larger sample to achieve sufficient statistical power to reliably distinguish the smaller effects of moderating variables (Cohen, 1992).

Effective qualitative research to better understand the motivation of workers/associates, *specifically in relation to improvement activity*, is a necessary prerequisite to a larger study. The assumed influence that organisational-level factors such as effective leadership have on individual motivation also requires further scrutiny as it applies to the 'HR Capability' → 'Lean Actions' → 'Effectiveness/Success' cause-effect relationship. At this point, many potential research questions remain — does the 'general' motivation of individuals affect their ability to carry out effective QI? How does a goal-specific motivation alter this underlying motivation? Healthcare is often delivered in environments of high stress, coupled with funding constraints that affect staff numbers, remuneration and workplace facilities. Although an unscientific observation, the researcher notes a very wide range of human ability to flourish under these circumstances. Clearly some individuals adapt better to carry out their work in these difficult conditions, and this work often includes the ability to participate in and deliver QI. Do some individuals simply have a higher ability to remain motivated/engaged, or can nominally unmotivated or disengaged staff still successfully complete a project if they believe in it strongly enough? Is goal setting theory, or something close to it, the best proposed fit? Properly examined, with sufficient statistical power, is the hypothesis that motivation is a moderating variable (with only weak effects on QI effectiveness) convincing enough to suggest practical efforts should therefore turn to the already demonstrated causal influence of HR Capability — as shown by this study and reported by other researchers (Hoegl & Gemuenden, 2001; Jayamaha et al., 2014; Kaplan et al., 2013; Lemieux-Charles et al., 2002).

The 'HR Capability' contextual factors were still treated at a high level of abstraction in this study, and examining the relative influence of the sub-components within the 'Teamwork' and 'Respect for People' constructs in much more detail is necessary. For example, exploring which elements of the Teamwork Quality Construct proposed by Hoegl and Gemuenden (2001) — *communications, cohesion, effort, mutual support, balance of contributions or coordination* — have the greatest effect and why. At present this remains guesswork for QI practitioners. Are formal communication mechanisms such as meetings and reports more important than informal conversations, arising from accessibility to colleagues and team cohesion? How is the quality of the team *interactions,* as considered by this construct, in turn affected by team make up and individual task skill?

The ‘Respect for People’ variable will also benefit from further analysis in terms of comparing contextual factors from multiple source models. This study aligned ‘Respect for People’ from the Toyota Way (TW) with team norms, principles and behaviours. These are of course important, but as Rother (2010) notes, there are differences in how ‘respect’ is interpreted in different cultures. ‘Respect for People’ also has a more specific meaning within the TW beyond simply ‘being respectful’ i.e. maintaining open dialog, providing the necessary support to achieve outcomes, coaching and investing in human potential (Jayamaha et al., 2014; Rother 2010). This very specific development of people focus is present, but somewhat buried in the MUSIQ model language of *QI leadership*, *QI culture* and *QI capability*.

Revisiting the ‘HR Capability’ → ‘Lean Actions’ → ‘Effectiveness/Success’ path at the sub-component level will also allow researchers to consider relationships beyond the strictly linear interaction path. The finding of Lemieux-Charles et al., (2002) that QI improves teamwork is worthy of more exploration, i.e. a reinforcing feedback loop is created, where teams carrying out QI improve their teamwork alongside the specific QI objectives. They are then better placed to generate and manage subsequent QI activity. Intuitively, this seems like a manifestation of the Toyota improvement and coaching ‘Kata’ or routines, an essential feature of the TW “Respect for People’ pillar (Rother, 2010).

From a practical research perspective, this topic does at least seem able to be studied, with some care. Even if the research continues to remain confined to the NZ secondary healthcare setting, there are many potential ‘teams’ able to be identified and studied, via qualitative and quantitative approaches. However, expanding the scope should be considered. The literature review for the present study revealed significant commonality of CSFs across industry sectors, and future research would be enhanced by studying beyond healthcare. This would mean that any tool developed for assessing the likelihood of QI effectiveness/success will be more widely applicable across different organisational settings. In addition, research that compares healthcare QI alongside QI in other sectors (rather than being confined to a healthcare only scope), will help identify those areas, if any, where healthcare QI ought to be treated as a different domain, requiring uniquely developed QI strategies and methods.

8. References

American Society of Manufacturing Engineers. (2008*). Lean certification body of knowledge version 3.0.* Available from http://www.sme.org/lean-certification.aspx

Andersen, H., Røvik, K. A., & Ingebrigsten, T. (2014). Lean thinking in hospitals: is there a cure for the absence of evidence? A systematic review of reviews, *BMJ Open* 2014; 4:e003873. Doi:10.1136/bmjopen-2013-003873

Anderson, T.W., & Darling, D.A. (1954). A Test of Goodness of Fit. *Journal of the American Statistical Association*, 49 (268), 765-769.

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*, 1173-1182.

Barson, S., Doolan-Noble, F., Gray , J., & Gauld, R. (2017). Healthcare leaders' views on successful quality improvement initiatives and context *Journal of Health Organization and Management*, Vol. 31 Iss. 1.

Borkowski, N. (2016). *Organizational behavior in health care*. 13th Ed. Burlington, Massachusetts : Jones & Bartlett Learning.

Bortolotti, T., Boscari, S., & Danese, P. (2015). Successful lean implementation: Organizational culture and soft lean practices, *International Journal of Production Economics*, Volume 160, February 2015, Pages 182-201, ISSN 0925-5273, http://doi.org/10.1016/j.ijpe.2014.10.013

Brandão de Souza, L. & Pidd, M. (2011). Exploring the barriers to lean health care implementation, *Public Money & Management*, 31:1, 59-66, DOI: 10.1080/09540962.2011.545548

Brennan, S.E., Bosch, M., Buchan, H., Green, S.E. (2012). Measuring organisational and individual factors thought to influence the success of quality improvement in primary care: a systematic review of instruments. *Implementation Science*, 7:121.

Bryman, A. M. (2012). *Social Research Methods* (4th Ed.). Oxford: Oxford University Press.

Burns, R. B. (2000). Ethics of Research, chapter 2 in *Introduction to Research Methods*, Longman Publishing.

Byrne, B. M. (2010). *Structural equation modeling with AMOS: basic concepts, applications, and programming* (2 ed. Vol. 2). NY: Routledge.

Campbell, D. E., Parks, C., & Wells, J. D. (2015). Understanding the Role of Theory on Instrument Development: An Examination of Strengths and Weaknesses of Discriminant Validity Analysis Techniques. *JITTA: Journal of Information Technology Theory and Application*, *16*(1), 25.

Canterbury District Health Board. (2017). *Canterbury DHB Annual Plan 2016/2017.* http://www.cdhb.health.nz/About-CDHB/corporate-publications/Documents/Canterbury%20Annual%20Plan%202016-2017.pdf [Retrieved 20 June 2017].

Chin, W. W. (1998). The partial least squares approach to structural equation modeling. In G. A. Marcoulides (Ed.), *Modern Methods for Business Research* (pp. 295-336). Mahwah, NJ: Lawrence Erlbaum Associates Inc.

Cohen, J. (1992). A power primer. Psychological bulletin, 112(1), 155-159.

Collins, L., Graham, J., & Flaherty, B. (1998). An alternative framework for defining mediation. *Multivariate Behavioral Research*, *33*, 295-312.

Crawley, M. J. (2015). *Statistics: an introduction using R.* (2nd Ed.). London: Wiley.

Creswell, J. W. (2002). *Educational research: Planning, conducting, and evaluating quantitative* (pp. 146-166). Upper Saddle River, NJ: Prentice Hall.

Creswell, J. W., & Poth, C. N. (2017). *Qualitative inquiry and research design: Choosing among five approaches*. Thousand Oaks, CA: Sage.

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

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52, 281-302.

Damij, N., Levnajić, Z., Skrt, V. R., & Suklan, J. (2015). What motivates us for work? Intricate web of factors beyond money and prestige. *PloS one*, *10*(7), e0132641.

Damschroder, L., Aron, D., Keith, R., Kirsh, S., Alexander, J., & Lowery, J. (2009). Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. *Implementation Science* 2009, 4:50 doi:10.1186/1748-5908-4-50

D'Andreamatteo, A., Ianni L., Lega F., & Sargiacomo M., (2015). Lean in Healthcare: a comprehensive review, *Health Policy* (2015), http://dx.doi.org/10.1016/j.healthpol.2015.02.002

DeVellis, R.F. (2012). *Scale development: theory and applications*. (3rd Ed.). Los Angeles: Sage.

Dolea, C., & Adams, O. (2005). Motivation of health care workers — review of theories and empirical evidence, *Cahiers de Sociologie et Demographie Medicale*. 2005 Jan-Mar;45(1):135-61.

Dubin, R. (1978). *Theory development*. New York: Free Press.

Emiliani, M. L. 2006. "Origins of Lean Management in America: The Role of Connecticut Businesses." *Journal of Management History* 12 (2): 167—184.10.1108/13552520610654069

Fornell, C., & Larcker, D. F. (1981). Evaluation of structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, 18(2), 39-50.

Gagné, M., & Deci, E. (2005) Self-determination theory and work motivation *Journal of Organizational Behavior* J. Organiz. Behav. 26, 331—362 .DOI: 10.1002/job.322

Gamage, P., Jayamaha, N. P., & Grigg, N. P. (2016). Acceptance of Taguchi's Quality Philosophy and Practice by Lean practitioners in apparel manufacturing. *Total Quality Management & Business Excellence*, 1-17.

Gonzalez, Aleu, F., & Van Aken, E. M. (2016). Systematic literature review of critical success factors for continuous improvement projects. *International Journal of Lean Six Sigma*, *7*(3), 214-232.

Graban, M. (2009). *Lean Hospitals: Improving quality, patient safety and employee satisfaction*. New York: CRC Press.

Griffin, A., McKeown, A., Viney, R., Rich, A., Welland, T., Gafson, I., & Woolf, K. (2017). Revalidation and quality assurance: the application of the MUSIQ framework in independent verification visits to healthcare organisations. *BMJ Open* 2017; 7:e014121.doi:10.1136/bmjopen-2016-014121

Grigg, N. P., & Jayamaha, N. P. (2014). Measuring quality and excellence: On the role of structural equation modelling in quality and business excellence research In P. K. Chopra (Ed.), *Quality, excellence and measurement: A tribute to Professor Gopal K. Kanji* (pp. 323-348). Leeds (UK): Wisdom House Publications Ltd.

Guba, E. G., & Lincoln, Y. S. (1994). Competing paradigms in qualitative research. *Handbook of qualitative research*, *2*(163-194), 105-117.

Guba, E. G. (Ed.). (1990). *The paradigm dialog*. Thousand Oaks, CA: Sage.

Hackman, J. R., & Oldham, G. R. (1976). Motivation through the design of work: Test of a theory. *Organizational behavior and human performance*, 16(2), 250-279.

Haenlein, M., & Kaplan, A. M. (2004). A beginner's guide to partial least squares analysis. *Understanding statistics*, *3*(4), 283-297.

Hair, J. F., Jr., , Anderson, R. E., Tatham, R. L., & Black, W. C. (1998). *Multivariate data analysis with readings* (5 ed.). Englewood Cliffs, NJ: Prentice Hall.

Hair, J. F., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2014). *A primer on partial least squares structural equation modeling (PLS-SEM)*. Calif.: Sage.

Harman, H. H. (1976). *Modern factor analysis*. University of Chicago Press.

Hoegl, M., & Gemuenden, H.G. (2001). Teamwork Quality and the Success of Innovative Projects: A Theoretical Concept and Empirical Evidence. *Organization Science* 12(4):435-449. http://dx.doi.org/10.1287/orsc.12.4.435.10635

Holweg, M. (2007). The genealogy of lean production. *Journal of operations management*, 25(2), 420-437.

Hovlid, E., & Bukve, O. (2014). A qualitative study of contextual factors' impact on measures to reduce surgery cancellations. *BMC Health Services Research* 2014, 14:215.

Iacobucci, D. (2010). Structural equations modeling: Fit indices, sample size, and advanced topics. *Journal of Consumer Psychology*, *20*(1), 90-98.

Ichijo, K., & Kohlbacher, F. (2008). Tapping tacit local knowledge in emerging markets—the Toyota way. *Knowledge Management Research & Practice*, 6(3), 173-186.

Jayamaha, N.P., Wagner, J.P., Grigg, N.P., Campbell-Allen, N.M., & Harvie, W. (2014). Testing a theoretical model underlying the 'Toyota Way' — an empirical study involving a large sample of Toyota facilities, *International Journal of Production Research*, 52:14, 4332-4350, DOI: 1080/00207543.2014.883467

Joosten, T., Bongers, I., & Janssen, R. (2009). Application of lean thinking to health care: issues and observations. *International Journal for Quality in Health Care*, 21(5), 341–347. http://doi.org/10.1093/intqhc/mzp036

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measuremen*t, 20, 141-151.

Kaplan, G.S., Patterson, S.H., Ching, J.M., Blackmore, C.C. (2014). Why Lean doesn't work for everyone *BMJ Qual Saf* Published Online First: 23 July 2014. doi: 10.1136/bmjqs-2014-003248

Kaplan, H.C., Brady, P.W., Dritz, M. C., Hooper, D. K., Linam, W. M., Froehle, C. M., & Margolis, P. (2010).The influence of context on quality improvement success in health care: a systematic review of the literature. *The Milbank Quarterly*, 88, 4, 500—559.

Kaplan, H. C., Provost, L. P., Froehle, C. M., & Margolis, P. A. (2012). The model for understanding success in quality (MUSIQ): building a theory of context in healthcare quality improvement. *BMJ Qual Saf*, 21, 13-20.

Kaplan, H. C., Froehle, C. M., Cassedy, A., Provost, L. P., & Margolis, P. A. (2013). An exploratory analysis of the model for understanding success in quality. *Health Care Management Review*, 2013, 38, 4, 325-338.

Kline, R. B. (2011). *Principles and practice of structural equation modeling*. New York, NY: Guilford press.

Kollberg, B., Dahlgaard, J. J., & Brehmer, P. O. (2006). Measuring lean initiatives in health care services: issues and findings. *International Journal of Productivity and Performance Management*, *56*(1), 7-24.

Lemieux-Charles, L., Murray, M., Baker, G.R., Barnsley, J., Tasa, K., & Ibrahi, S.A., (2002). The effects of quality improvement practices on team effectiveness: a mediational model. Journal *of Organizational Behavior J. Organiz. Behav.* 23, 533—553 (2002) Published online 17 May 2002 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/job.154

Liker, J. (2004). *The Toyota Way: 14 management principles from the world's greatest manufacturer*. New York; McGraw-Hill.

Liker, J. K., & Hoseus, M. (2009). Human resource development in Toyota culture. *International Journal of Human Resources Development and Management*, 10(1), 34-50.

Lindner, J. R. (1998). Understanding employee motivation. *Journal of extension*, *36*(3), 1-8.

Luthans, F., Luthans, B.C., & Luthans, K.W. (2015*). Organizational behavior: an evidence-based approach.* 13th Ed. Charlotte, North Carolina : Information Age Publishing, Inc.

Machado Guimarães, C., & Crespo de Carvalho, J. (2014). Assessing Lean Deployment in Healthcare—A Critical Review and Framework, *Journal of Enterprise Transformation*, 4:1, 3-27, DOI: 10.1080/19488289.2013.869277

Maher, L., Gustafson, D., and Evans, A. (2008). *The NHS Sustainability Model.* http://webarchive.nationalarchives.gov.uk/20150401105500/http://www.institute.nhs.uk/sustainability_model/introduction/find_out_more_about_the_model.html [Retrieved 14 April 2017].

Malterud, K. (2001). The art and science of clinical knowledge: evidence beyond measures and numbers. *The Lancet*, *358*(9279), 397-400.

May, D. R., Gilson, R. L., & Harter, L. M. (2004). The psychological conditions of meaningfulness, safety and availability and the engagement of the human spirit at work. *Journal of occupational and organizational psychology*, 77(1), 11-37.

Mays, N. (2013). *Reorienting the New Zealand health care system to meet the challenge of long-term conditions in a fiscally constrained environment.* Revised version of paper prepared for New Zealand Treasury Long-term Fiscal External Panel, November 2012, and Chair of Public Finance, Victoria University of Wellington and New Zealand Treasury conference, Affording our Future, Wellington, 10-11 December 2012.

Marsden, P.V., & Wright, J. D. (2010). *Handbook of Survey Research.* (2nd Ed.). Bingley, UK: Emerald Group Publishing.

Massey University. (2015). *Code of Ethical Conduct for Research, Teaching and Evaluations involving Human Participants.*

Messick, S. (1995). Validity of psychological assessment: validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50(9), 741-749.

Mevik, B. H., & Wehrens, R. (2007). The pls package: principal component and partial least squares regression in R. *Journal of Statistical software*, *18*(2), 1-24

Ministry of Health. (2012*). Improving the System: Meeting the challenge — improving patient flow for electives. A Toolkit for District Health Boards.* Wellington: Ministry of Health.

Ministry of Health. (2014). *Briefing to the incoming minister 2014*. Wellington: Ministry of Health. www.health.govt.nz [Retrieved April 15 2015].

Ministry of Health. (2017). *Overview of the health system.* http://www.health.govt.nz/new-zealand-health-system/overview-health-system [Retrieved 06 June 2017].

Mora, M. (2011). *Why We Need to Avoid Long Surveys.* https://www.relevantinsights.com/long-surveys/ [Retrieved 17 April 2017].

Moraros, J., Lemstra, M., & Nwankwo, C. (2016). Lean interventions in healthcare: do they actually work? A systematic literature review. *International Journal for Quality in Health Care*, 2016, 28(2), 150-165 DOI: 10.1093/intqhc/mzv123

Morgan, G. (2007). *Images of organization*, Thousand Oaks, CA: Sage.

Mulaik, S. A. (2009). *Linear causal modeling with structural relations*. NW: CRC Press.

National Institute of Standards and Technology (2017) https://www.nist.gov/baldrige/baldrige-faqs-baldrige-award-recipients)

Nelson, E., Godfrey, M., Batalden, P., Berry, S., Bothe, A., McKinley, K., Melin, C., Muething, S., Moore, L.G., Wasson, J., Nolan, T., (2008). Clinical Microsystems, Part 1. The Building Blocks of Health Systems. *The Joint Commission Journal on Quality and Patient Safety*. Volume 34, Number 7.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.): New York: McGraw-Hill.

Ohno, T. (1988). *Toyota Production System: Beyond Large-scale Production*. Portland, ME: Productivity Press.

Øvretveit, J. (2010). Understanding the conditions for improvement: research to discover which context influences affect improvement success. *BMJ Qual Saf* 2011, 20 (Suppl. 1):i18ei23. doi:10.1136/bmjqs.2010

Podsakoff, P.M., MacKenzie, S.B., Lee, J.Y., & Podsakoff, N.P. (2003). Common method biases in behavioral research: a critical review of the literature and recommended remedies. *Journal of Applied Psychology* (88:5) 2003, p.879.

Portela, M.C., Pronovost, P.J., Woodcock, T., Carter, P., & Dixon-Woods, M. (2015). How to study improvement interventions: a brief overview of possible study types. *BMJ Qual Saf*. 2015, 24, 325—336.

Philliber, S.G., Schwab, M.R. & Sloss, G.S. (1980). “Study Design,” Chapter 4 in: *Social Research*, F.E. Peacock Publishers, Itasca Illinois.

R Core Team (2014). R: *A Language and Environment for Statistical Computing*, R Foundation for Statistical computing, Vienna. Available from http://ww.r-project.org/

Ringle, C. M., Wende, S., and Becker, J.-M. (2015). "SmartPLS 3." Boenningstedt: SmartPLS GmbH, http://www.smartpls.com

Rother, M. (2010). *Toyota kata*. New York: MacGraw Hill.

Rubenstein, L., Khodyakov, D., Hempel, S., Danz, M., Salem-Schatz, S., Foy, R., O’Neill, S., Dalal, S., & Shekelle, P. (2014). How can we recognize continuous quality improvement? *International Journal for Quality in Health Care*, 26, 1, 6—15.

Sampaio, P., Saraiva, P., and Monteiro, A. (2012). "A comparison and usage overview of business excellence models", *The TQM Journal*, Vol. 24 Issue: 2, pp.181-200, doi: 10.1108/17542731211215125

Saurin, T., Rooke, J., & Koskela, L. (2013). A complex systems theory perspective of lean production, *International Journal of Production Research*, 51:19, 5824-5838, DOI: 10.1080/00207543.2013.796420

Schouten, L., Grol, R., & Hulscher, M. (2010). Factors influencing success in quality-improvement collaboratives: development and psychometric testing of an instrument. *Implementation Science* 2010, 5:84.

Seo, M., Barrett, L., and Bartunek, J. (2004). The role of affective experience in work motivation. *The Academy of Management Review*, Vol. 29, No. 3 (2004), pp. 423-439.

Shah, R. & Ward, P.T. (2007). Defining and developing measures of lean production. *Journal of Operations Management*. 25(4), 785-805.

Shalley, C. E., & Gilson, L. L. (2004). What leaders need to know: A review of social and contextual factors that can foster or hinder creativity. *The leadership quarterly*, 15(1), 33-53.

Shojania, K. G., & Grimshaw, J. M. (2005). Evidence-based quality improvement: the state of the science, *Health Affairs*, 24, 1, 138-150.

Sisson, J., & Elshennawy, A. (2015)," Achieving success with Lean: An analysis of key factors in Lean transformation at Toyota and beyond ", *International Journal of Lean Six Sigma*, Vol. 6 Iss. 3 pp. 263 - 280.

Sreedharan, V., and R. Raju, R. (2016)."A systematic literature review of Lean Six Sigma in different industries ", *International Journal of Lean Six Sigma*, Vol. 7 Iss. 4 pp. 430 - 466.

Stajkovic, A. D., & Luthans, F. (2001). Differential effects of incentive motivators on work performance. Academy of management journal, 44(3), 580-590.

Statistics New Zealand. (2013). *2013 Census district health board tables* http://www.stats.govt.nz/Census/2013-census/data-tables/dhb-tables.aspx [Retrieved 06 June 2017].

Straub, D., Boudreau, M., & Gefen, D. (2004). Validation guidelines for IS positivist research. *Communications of the Association for Information Systems*, 13, 380-427.

Taylor, M. J., McNicholas, C., Nicolay, C., Darzi, A., Bell, D., & Reed, J. E. (2013). Systematic review of the application of the plan–do–study–act method to improve quality in healthcare. *BMJ quality & safety*, 0:1–9. doi:10.1136/bmjqs-2013-001862

Walshe, K. (2007). Understanding what works—and why—in quality improvement: the need for theory-driven evaluation. *International Journal for Quality in Health Care*, 19, 2, 57—59.

Waring, J., & Bishop, S. (2010). Lean healthcare: rhetoric, ritual and resistance, *Social Science and Medicine*, 71, 1332-1340.

White, M. (2015). How effective is productive ward? *Nursing Times*, 111, 8, 22-25.

Womack, J. P., & Jones, D. T. (2003). *Lean thinking: banish waste and create wealth in your corporation.* (2nd Ed.). London: Simon and Schuster.

Womack, J. P., & Jones, D. T. (2005). Lean consumption. *Harvard business review*, *83*(3), 58-68.

Wong, K. K. K. (2013). Partial least squares structural equation modeling (PLS-SEM) techniques using SmartPLS. Marketing Bulletin, 24(1), 1-32.

Worchel, S. (1986). The influence of contextual variables on interpersonal spacing. *Journal of Nonverbal Behavior*, 10(4), 230-254. doi:10.1007/bf00987482

World Health Organisation. (2000). *The world health report 2000 - Health systems: improving performance.* http://www.who.int/whr/2000/en/ [Retrieved 06 June 2017].

Yilmaz, K. (2013). Comparison of quantitative and qualitative research traditions: Epistemological, theoretical, and methodological differences. *European Journal of Education*, *48*(2), 311-325.

Yin, R. K. (2013). Case study research: Design and methods. Los Angeles: Sage.

Young, T.P., & McClean, S.I. (2008). A critical look at lean thinking in healthcare, *BMJ Quality and Safety, Qual Saf Health Care*, 17, 382-386.

Zammito, J. H. (2004). *A nice derangement of epistemes: Post-positivism in the study of science from Quine to Latour*. Chicago, TX: University of Chicago Press.

Appendices

Appendix 1: Complete List of Survey Items

Questions for 2nd cohort of respondents (post pilot; Collabor8 and RT2C OPH full) April-May 2016	Source of Items**	Retained for PLSPM
1. Our initiative identified waste in work processes.	Original	Yes
2. Our initiative identified value-adding activity in work processes.	Original	Yes
3. Our initiative attempted to identify the underlying causes of process problems.	Original	Yes
4. Our initiative observed operational staff in their workplace.	Original	Yes
5. Our initiative used visual tools as part of the solution to support operational processes.	Original	Yes
6. Our initiative attempted to develop safe, reliable and efficient procedures for staff to follow	Original	Yes
7. Our initiative used a Plan-Do-Study-Act framework.	Original	Yes
8. Each member of our team understood their role in the initiative.	Schouten et al. (2010)	No
9. We maintained a focused effort.	Hoegl & Gemuenden (2001)	Yes
10. Our team included experts in the subject area we were trying to improve.	Schouten et al. (2010)	No
11. Team members took responsibility for their actions and behaviours.	Hoegl & Gemuenden (2001)	Yes
12. We avoided personal criticism when reviewing ideas within the team.	Lemieux-Charles et al. (2002)	Yes
13. We resolved conflicts respectfully.	Lemieux-Charles et al. (2002)	No
14. I enjoyed being part of this team.	Original	Yes
15. Team members supported each other.	Hoegl & Gemuenden (2001)	Yes
16. We had clear goals.	Lemieux-Charles et al., (2002)	Yes
17. We had regular meetings.	Hoegl & Gemuenden (2001)	Yes
18. We used facts and data to understand performance.	Lemieux-Charles et al. (2002)	Yes
19. We knew how we were progressing towards our goals.	Lemieux-Charles et al. (2002)	Yes
20. All of our team were encouraged to contribute ideas.	Hoegl & Gemuenden (2001); Lemieux-Charles et al. (2002)	Yes
21. I believed this initiative was an appropriate use of our time and effort.	Lemieux-Charles et al. (2002)	Yes
22. I believed there were better ways to improve the focus area than the solution we chose.	Lemieux-Charles et al. (2002)	Yes
23. I was worried about what might go wrong.	Original	No
24. I valued this initiative but was distracted by other work concerns.	Original	No
25. Did you experience any setbacks during the project?	Original	No
26. We were able to learn from our setbacks.	Original	No
27. Our goals were unrealistic.	Original	Yes
28. We felt fully supported by management.	Original	Yes
29. I valued this initiative but was distracted by non-work concerns.	Original	No
30. I felt that I was being made to do this improvement initiative despite my concerns.	Original	No

Questions for 2nd cohort of respondents (post pilot; Collabor8 and RT2C OPH full) April-May 2016	Source of Items**	Retained for PLSPM
31. I believed in this initiative, despite not knowing how to achieve successful outcomes.	Original	No
32. I knew why we were attempting this improvement initiative.	Original	Yes
33. How many improvement initiatives had you participated in prior to this one?	Original	No
34. Have you ever studied quality improvement methods and techniques?	Original	No
35. How many quality improvement training courses or workshops have you attended?	Original	No
36. How many Lean training courses or workshops have you attended?	Original	No
37. How many different work environments have you worked in over the last 10 years?	Original	No
38. Our initiative was successful in meeting its stated aim.	Kaplan et al. (2013)	Yes
39. Our initiative achieved positive outcomes for patients, staff or the health system.	Original	Yes
40. Our initiative achieved additional, unexpected benefits.	Original	No
41. Please assign a score between 0-100 for your assessment of the success of your initiative.	Kaplan et al. (2013)	Yes
42. Are there any data or measurements associated with your initiative?	N/A	N/A
43. Are there any reports or stories associated with your initiative?	N/A	N/A
44. Having taken some time to reflect on your initiative during this survey, what one thing would you do differently?	N/A	N/A
45. What were the primary objectives of your initiative?	N/A	N/A
46. Was your initiative associated with a formal improvement programme?	N/A	N/A
47. Is your initiative completed or still in progress?	N/A	N/A
48. How many people were in your team?	N/A	N/A
49. Was your personal role in the initiative primarily operational or as an external contributor?	N/A	N/A
50. What was the role mix in your team? [Senior medical officer]	N/A	N/A
50. What was the role mix in your team? [Medical officer]	N/A	N/A
50. What was the role mix in your team? [Nurse]	N/A	N/A
50. What was the role mix in your team? [Hospital aide]	N/A	N/A
50. What was the role mix in your team? [Allied health practitioner]	N/A	N/A
50. What was the role mix in your team? [Quality facilitator]	N/A	N/A
50. What was the role mix in your team? [Project manager/Project lead]	N/A	N/A
50. What was the role mix in your team? [Administrator]	N/A	N/A
50. What was the role mix in your team? [Manager]	N/A	N/A
50. What was the role mix in your team? [Analyst]	N/A	N/A
50. What was the role mix in your team? [Other]	N/A	N/A
51. Location of your improvement initiative.	N/A	N/A
52. Would you be willing to participate in a short follow-up interview?	N/A	N/A
53. Contact email	N/A	N/A

*** 'Source' refers to an existing survey instrument used to identify an item of similar function or intent, not an exact replica of the wording*

This version is the master list for survey question numbers
Questions in red text were reverse scored at time of coding for the first iteration of PLSPM. Original values were restored for final iteration and hypothesis test

Questions in amber text were not intended for the statistical analysis, but were included for context and additional information of potential CDHB interest

Appendix 2: Bivariate Correlation Analysis

Construct	Lean Actions							Teamwork						Respect for People				Motivation					Experience		Perceived Success		
Question	I1	I2	I3	I4	I5	I6	I7	I9	I16	I17	I18	I19	I20	I11	I12	I14	I15	I21	I28	I32	I22	I27	I33	I34	I38	I39	I41
I1	1.000	0.418	0.463	0.198	0.294	0.372	0.411	0.261	0.281	0.254	0.461	0.370	0.301	0.101	0.172	0.209	0.059	0.409	0.320	0.338	0.279	0.189	0.091	0.134	0.372	0.486	0.437
I2	0.418	1.000	0.379	0.180	0.337	0.351	0.303	0.331	0.197	0.320	0.407	0.306	0.162	0.213	0.172	0.347	0.044	0.529	0.307	0.488	0.231	0.135	0.174	0.106	0.417	0.496	0.442
I3	0.463	0.379	1.000	0.352	0.258	0.307	0.261	0.118	0.215	0.219	0.474	0.143	0.254	0.208	0.149	0.249	0.117	0.440	0.257	0.440	0.208	0.125	-0.090	0.069	0.243	0.334	0.210
I4	0.198	0.180	0.352	1.000	0.486	0.492	0.298	0.155	0.262	0.261	0.506	0.144	0.256	0.119	0.158	0.213	0.077	0.197	0.266	0.400	0.021	0.222	0.080	0.190	0.215	0.319	0.280
I5	0.294	0.337	0.258	0.486	1.000	0.394	0.381	0.243	0.291	0.464	0.463	0.279	0.390	0.248	0.180	0.363	0.188	0.324	0.323	0.378	0.096	0.102	0.209	0.131	0.372	0.383	0.274
I6	0.372	0.351	0.307	0.492	0.394	1.000	0.586	0.169	0.256	0.263	0.507	0.293	0.250	0.068	0.043	0.248	0.000	0.413	0.173	0.411	0.246	0.281	0.196	0.097	0.375	0.497	0.464
I7	0.411	0.303	0.261	0.298	0.381	0.586	1.000	0.306	0.322	0.311	0.392	0.339	0.403	0.013	0.010	0.255	0.055	0.393	0.266	0.332	0.187	0.180	0.154	0.191	0.338	0.350	0.408
I9	0.261	0.331	0.118	0.155	0.243	0.169	0.306	1.000	0.422	0.384	0.326	0.508	0.420	0.547	0.446	0.454	0.501	0.348	0.323	0.207	0.432	0.239	-0.041	0.055	0.412	0.294	0.367
I16	0.281	0.197	0.215	0.262	0.291	0.256	0.322	0.422	1.000	0.242	0.509	0.509	0.453	0.322	0.302	0.461	0.387	0.455	0.406	0.484	0.373	0.500	0.080	0.166	0.530	0.500	0.421
I17	0.254	0.320	0.219	0.261	0.464	0.263	0.311	0.384	0.242	1.000	0.454	0.431	0.363	0.390	0.273	0.347	0.241	0.302	0.367	0.342	0.191	0.106	0.152	0.280	0.284	0.253	0.194
I18	0.461	0.407	0.474	0.506	0.463	0.507	0.392	0.326	0.509	0.454	1.000	0.514	0.485	0.247	0.235	0.372	0.230	0.536	0.467	0.592	0.403	0.352	0.053	0.213	0.483	0.588	0.498
I19	0.370	0.306	0.143	0.144	0.279	0.293	0.339	0.508	0.509	0.431	0.514	1.000	0.431	0.259	0.222	0.338	0.213	0.420	0.500	0.301	0.419	0.420	0.048	0.132	0.556	0.566	0.500
I20	0.301	0.162	0.254	0.256	0.390	0.250	0.403	0.420	0.453	0.363	0.485	0.431	1.000	0.300	0.382	0.440	0.294	0.413	0.385	0.430	0.315	0.249	0.008	0.126	0.340	0.314	0.249
I11	0.101	0.213	0.208	0.119	0.248	0.068	0.013	0.547	0.322	0.390	0.247	0.259	0.300	1.000	0.594	0.447	0.702	0.204	0.297	0.165	0.312	0.154	0.092	0.071	0.336	0.291	0.184
I12	0.172	0.172	0.149	0.158	0.180	0.043	0.010	0.446	0.302	0.273	0.235	0.222	0.382	0.594	1.000	0.633	0.585	0.269	0.196	0.029	0.292	0.092	0.103	0.085	0.229	0.208	0.206
I14	0.209	0.347	0.249	0.213	0.363	0.248	0.255	0.454	0.461	0.347	0.372	0.338	0.440	0.447	0.633	1.000	0.546	0.545	0.311	0.279	0.401	0.187	0.062	0.121	0.384	0.348	0.413
I15	0.059	0.044	0.117	0.077	0.188	0.000	0.055	0.501	0.387	0.241	0.230	0.213	0.294	0.702	0.585	0.546	1.000	0.291	0.340	0.136	0.373	0.190	-0.079	0.012	0.267	0.245	0.154
I21	0.409	0.529	0.440	0.197	0.324	0.413	0.393	0.348	0.455	0.302	0.536	0.420	0.413	0.204	0.269	0.545	0.291	1.000	0.436	0.614	0.426	0.431	0.058	0.129	0.552	0.590	0.563
I28	0.320	0.307	0.257	0.266	0.323	0.173	0.266	0.323	0.406	0.367	0.467	0.500	0.385	0.297	0.196	0.311	0.340	0.436	1.000	0.349	0.325	0.183	0.078	0.189	0.308	0.475	0.309
I32	0.338	0.488	0.440	0.400	0.378	0.411	0.332	0.207	0.484	0.342	0.592	0.301	0.430	0.165	0.029	0.279	0.136	0.614	0.349	1.000	0.264	0.433	0.151	0.177	0.512	0.566	0.434
I22	0.279	0.231	0.208	0.021	0.096	0.246	0.187	0.432	0.373	0.191	0.403	0.419	0.315	0.312	0.292	0.401	0.373	0.426	0.325	0.264	1.000	0.520	-0.039	0.016	0.349	0.438	0.432
I27	0.189	0.135	0.125	0.222	0.102	0.281	0.180	0.239	0.500	0.106	0.352	0.420	0.249	0.154	0.092	0.187	0.190	0.431	0.183	0.433	0.520	1.000	-0.025	0.088	0.449	0.502	0.451
I33	0.091	0.174	-0.090	0.080	0.209	0.196	0.154	-0.041	0.080	0.152	0.053	0.048	0.008	0.092	0.103	0.062	-0.079	0.058	0.078	0.151	-0.039	-0.025	1.000	0.354	0.194	0.120	0.158
I34	0.134	0.106	0.069	0.190	0.131	0.097	0.191	0.055	0.166	0.280	0.213	0.132	0.126	0.071	0.085	0.121	0.012	0.129	0.189	0.177	0.016	0.088	0.354	1.000	0.072	0.132	0.006
I38	0.372	0.417	0.243	0.215	0.372	0.375	0.338	0.412	0.530	0.284	0.483	0.556	0.340	0.336	0.229	0.384	0.267	0.552	0.308	0.512	0.349	0.449	0.194	0.072	1.000	0.761	0.656
I39	0.486	0.496	0.334	0.319	0.383	0.497	0.350	0.294	0.500	0.253	0.588	0.566	0.314	0.291	0.208	0.348	0.245	0.590	0.475	0.566	0.438	0.502	0.120	0.132	0.761	1.000	0.630
I41	0.437	0.442	0.210	0.280	0.274	0.464	0.408	0.367	0.421	0.194	0.498	0.500	0.249	0.184	0.206	0.413	0.154	0.563	0.309	0.434	0.432	0.451	0.158	0.006	0.656	0.630	1.000

Appendix 3: Principal Components Analysis

Full survey response set of 105; missing data imputed, n=105

Tests carried out using R Studio software (R Core Team, 2014).

The criteria for retaining components were Eigenvalues greater than 1. Criterion for retaining an indicator was a rotated loading to the component of at least 0.7.

Response variable: "Perceived Success"

Scale items:

1. Q38. Our initiative was successful in meeting its stated aim.
2. Q39. Our initiative achieved positive outcomes for patients, staff or the health system.
3. Q41. Please assign a score between 0-100 for your assessment of the success of your initiative.

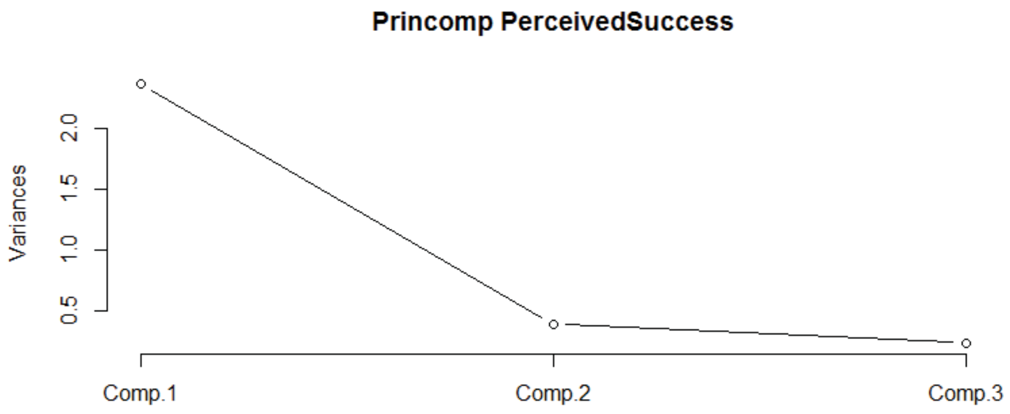
Scale reliability using Cronbach's alpha: 0.87

Details:

```
Reliability analysis
Call: alpha(x = PerceivedSuccess)

  raw_alpha std.alpha G6(smc) average_r S/N   ase mean   sd
      0.86      0.87    0.82      0.68 6.4 0.079  3.8 0.79

 lower alpha upper     95% confidence boundaries
0.7 0.86 1.01

 Reliability if an item is dropped:
    raw_alpha std.alpha G6(smc) average_r S/N alpha se
I38      0.76      0.77    0.63      0.63 3.4     0.14
I39      0.79      0.79    0.66      0.66 3.8     0.13
I41      0.86      0.86    0.76      0.76 6.4     0.12
```

Principal component analysis

Scree plot for Perceived Success principal components:



Detail:

```
Importance of components:
                          Comp.1    Comp.2     Comp.3
Standard deviation     1.5382576 0.6297214 0.48704682
Eigenvalue             2.36       0.39        0.23
Proportion of Variance 0.7887455 0.1321830 0.07907154
Cumulative Proportion  0.7887455 0.9209285 1.00000000
```

```
> Unrotated loadings
```

```
Loadings:
    Comp.1 Comp.2 Comp.3
I38 -0.592 -0.322  0.739
I39 -0.585 -0.458 -0.669
I41 -0.554  0.828

               Comp.1 Comp.2 Comp.3
SS loadings     1.000  1.000  1.000
Proportion Var  0.333  0.333  0.333
Cumulative Var  0.333  0.667  1.000
```

```
> Rotated loadings
```

```
Loadings:
    PC1
I38 0.911
I39 0.900
I41 0.852

                 PC1
SS loadings    2.366
Proportion Var 0.789
```

Comments:

The Eigenvalue for principal component 1 is 2.36, and the Eigenvalue for principal component 2 is 0.39. The rotated loadings show strong loading to PC1, accounting for 79% of the variance. I interpret these combined results to mean that the three measured variables are reflecting one common latent variable.

Cause variable: "Lean actions"

Scale items:

1. Q1. Our initiative identified waste in work processes.
2. Q2. Our initiative identified value-adding activity in work processes.
3. Q3. Our initiative attempted to identify the underlying causes of process problems.
4. Q4. Our initiative observed operational staff in their workplace.
5. Q5. Our initiative used visual tools as part of the solution to support operational processes.
6. Q6. Our initiative attempted to develop safe, reliable and efficient procedures for staff to follow
7. Q7. Our initiative used a Plan-Do-Study-Act framework.

Scale reliability using Cronbach's alpha: 0.80

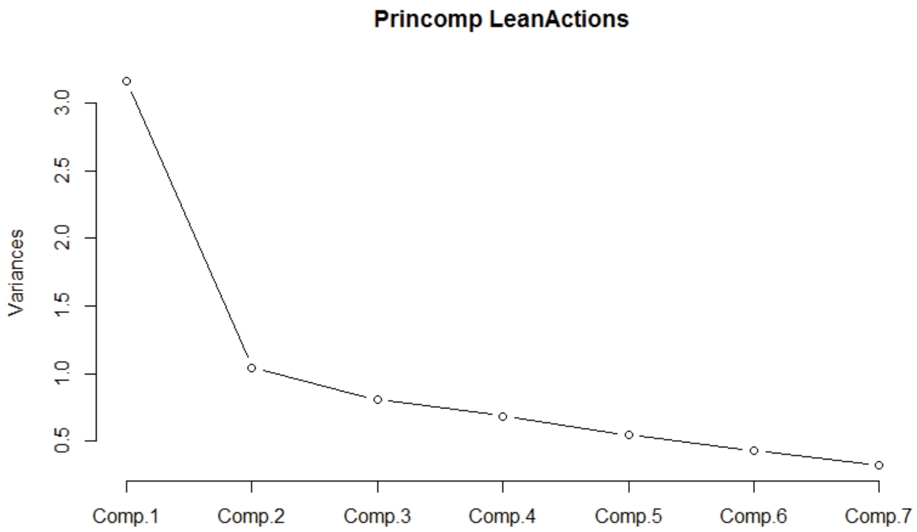
Details:

```
Call: alpha(x = LeanActions)

  raw_alpha std.alpha G6(smc) average_r S/N   ase mean   sd
      0.79       0.8      0.8      0.36 3.9 0.051  4.1 0.55

 lower alpha upper     95% confidence boundaries
0.69 0.79 0.89
```

Comments:

Principal component analysis

Scree plot for Lean actions principal components:



Details:

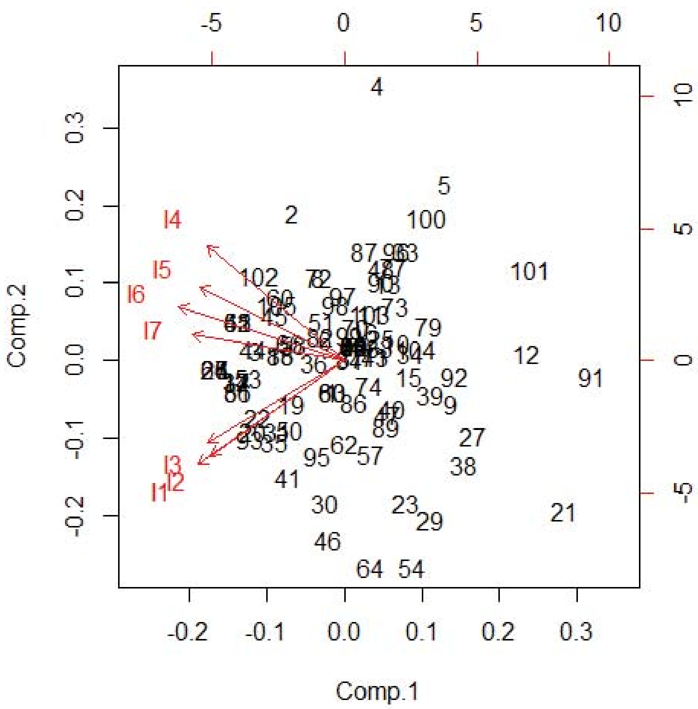
```
summary(fit)
Importance of components:
                          Comp.1    Comp.2    Comp.3     Comp.4     Comp.5     Comp.6
Standard deviation     1.7781518 1.0198335 0.8994008 0.82874444 0.73842353 0.65817982
Eigenvalue             3.16      1.04      0.808
Proportion of Variance 0.4516891 0.1485800 0.1155603 0.09811676 0.07789562 0.06188581
Cumulative Proportion  0.4516891 0.6002692 0.7158294 0.81394620 0.89184182 0.95372763

                           Comp.7
Standard deviation     0.56912793
Eigenvalue
Proportion of Variance 0.04627237
Cumulative Proportion  1.00000000
```

```
Unrotated Loadings:

Loadings:
   Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6 Comp.7
I1 -0.379 -0.470        -0.192  0.516  0.561
I2 -0.348 -0.432         0.629 -0.496         0.223
I3 -0.355 -0.373 -0.511 -0.415        -0.499 -0.213
I4 -0.356  0.519 -0.447 -0.181 -0.173  0.284  0.511
I5 -0.376  0.334 -0.206  0.551  0.523 -0.181 -0.308
I6 -0.430  0.244  0.329 -0.187 -0.407  0.269 -0.612
I7 -0.395  0.116  0.615 -0.155  0.114 -0.499  0.406

               Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6 Comp.7
SS loadings     1.000  1.000  1.000  1.000  1.000  1.000  1.000
Proportion Var  0.143  0.143  0.143  0.143  0.143  0.143  0.143
Cumulative Var  0.143  0.286  0.429  0.571  0.714  0.857  1.000
```

```
Rotated Loadings, 2 Principal Components:

Loadings:
   PC1   PC2
I1 0.181 0.806
I2 0.167 0.742
I3 0.216 0.705
I4 0.824
I5 0.725 0.193
I6 0.735 0.325
I7 0.602 0.382

                 PC1   PC2
SS loadings    2.215 1.987
Proportion Var 0.316 0.284
Cumulative Var 0.316 0.600
```

```
Rotated Loadings, 1 Principal Component:

Loadings:
   PC1
I1 0.673
I2 0.620
I3 0.632
I4 0.633
I5 0.669
I6 0.764
I7 0.703

              PC1
SS loadings    3.162
Proportion Var 0.452
```

Biplot



Moderating Variable 1: "Team Norms1" (functional interactions) Renamed as "Teamwork" in final model

Scale Items

1. Q9. We maintained a focused effort.
2. Q16. We had clear goals.
3. Q17. We had regular meetings.
4. Q18. We used facts and data to understand performance.
5. Q19. We knew how we were progressing towards our goals.
6. Q20. All of our team were encouraged to contribute ideas.

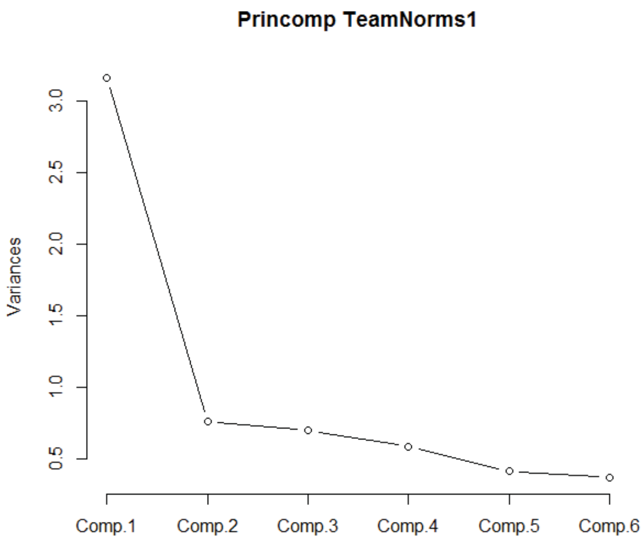
Scale reliability using Cronbach's alpha: 0.82

Details:

```
Reliability analysis
Call: alpha(x = TeamNorms1)

  raw_alpha std.alpha G6(smc) average_r S/N   ase mean   sd
      0.81      0.82    0.81      0.43 4.5 0.052  3.9 0.54

 lower alpha upper     95% confidence boundaries
0.71 0.81 0.92
```

Principal component analysis

Scree plot for Team Norms1 principal components:




```
Importance of components:
                          Comp.1    Comp.2    Comp.3     Comp.4     Comp.5     Comp.6
Standard deviation     1.7778280 0.8748430 0.8371385 0.76638249 0.64373860 0.60945474
Eigenvalues            3.16      0.76      0.7        0.58       0.41       0.37
Proportion of Variance 0.5267787 0.1275584 0.1168002 0.09789035 0.06906656 0.06190585
Cumulative Proportion  0.5267787 0.6543371 0.7711372 0.86902759 0.93809415 1.00000000
```

Varimax rotated loading to PC1

```
Loadings:
    PC1
I9  0.698
I16 0.726
I17 0.645
I18 0.763
I19 0.790
I20 0.725

                 PC1
SS loadings    3.162
Proportion Var 0.527
```

Comments:

The Eigenvalue for principal component 1 is 3.16 and the Eigenvalue for principal component 2 is 0.76. I interpret this result to mean that the measured variables are reflecting one latent construct.

Moderating Variable 2: "Team Norms2" (Interpersonal interactions) Renamed as Respect for people

Scale Items

1. Q11. Team members took responsibility for their actions and behaviours.
2. Q12. We avoided personal criticism when reviewing ideas within the team.
3. Q14. I enjoyed being part of this team.
4. Q15. Team members supported each other.

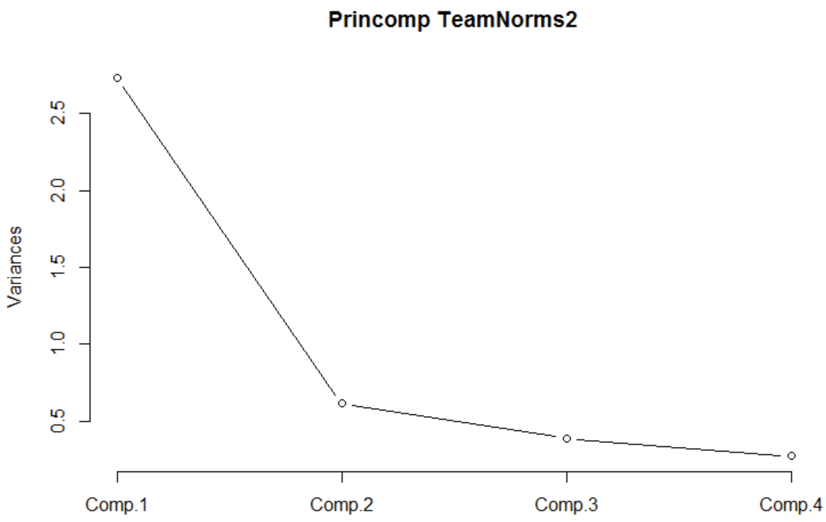
Scale reliability using Cronbach's alpha: 0.84

Details:

```
Reliability analysis
Call: alpha(x = TeamNorms2)

  raw_alpha std.alpha G6(smc) average_r S/N   ase mean   sd
      0.83      0.84    0.82      0.58 5.4 0.065  3.9 0.65

 lower alpha upper     95% confidence boundaries
0.71 0.83 0.96

 Reliability if an item is dropped:
    raw_alpha std.alpha G6(smc) average_r S/N alpha se
I11      0.79      0.80    0.73      0.57 4.0    0.088
I12      0.78      0.79    0.74      0.56 3.8    0.090
I14      0.82      0.83    0.77      0.62 4.9    0.085
I15      0.78      0.78    0.72      0.55 3.6    0.090
```

Principal component analysis

Scree plot for Team Norms 2 principal components:



Details:

```
Importance of components:
                          Comp.1    Comp.2    Comp.3     Comp.4
Standard deviation     1.6520898 0.7826309 0.6215954 0.52125558
Eigenvalues            2.72       0.61        0.38     0.27
Proportion of Variance 0.6823502 0.1531278 0.0965952 0.06792684
Cumulative Proportion  0.6823502 0.8354780 0.9320732 1.00000000
```

Varimax rotated loading to PC1

```
Loadings:
    PC1
I11 0.829
I12 0.848
I14 0.783
I15 0.857

                 PC1
SS loadings    2.756
Proportion Var 0.689
```

Comments:

The Eigenvalue for principal component 1 is 2.72 and the Eigenvalue for principal component 2 is 0.61. I interpret this result to mean that the measured variables are reflecting one latent construct.

Moderating Variable 3: "Motivation"

Note –

Corrected Scale Items (unreliable questions removed):

1. Q21. I believed this initiative was an appropriate use of our time and effort.
2. Q22. I believed there were better ways to improve the focus area than the solution we chose.
3. Q27. Our goals were unrealistic.
4. Q28. We felt fully supported by management.

5. Q32. I knew why we were attempting this improvement initiative.

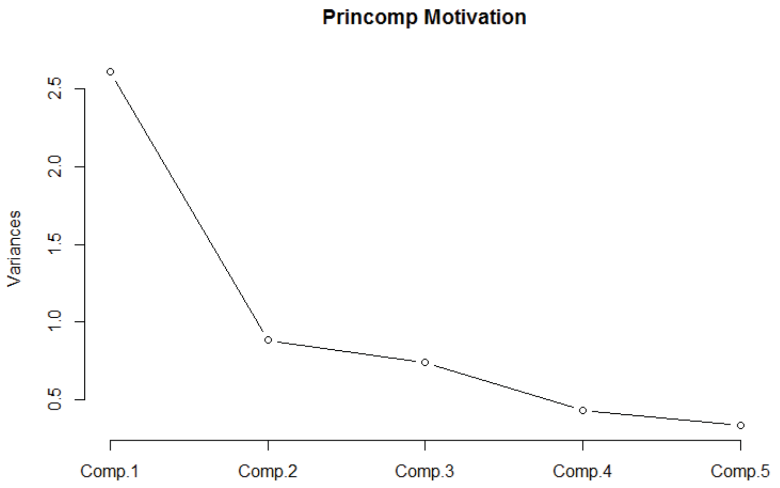
Corrected Scale reliability using Cronbach's alpha: 0.77

Details:

```
Reliability analysis
Call: alpha(x = Motivation)

 raw_alpha std.alpha G6(smc) average_r S/N   ase mean   sd
      0.76      0.77    0.76       0.4 3.3 0.065  3.8 0.61

 lower alpha upper     95% confidence boundaries
0.63 0.76 0.89
```

Principal component analysis

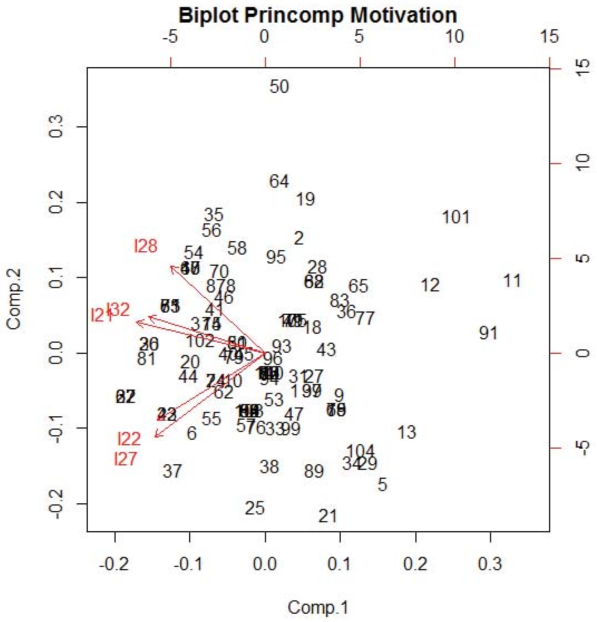
Scree plot for Motivation principal components:



Details:

```
Importance of components:
                          Comp.1    Comp.2    Comp.3     Comp.4     Co
mp.5
Standard deviation     1.6141286 0.9399102 0.8623022 0.65761942 0.5789
0350
Eigenvalues            2.60       0.88       0.74        0.43       0.33
Proportion of Variance 0.5210822 0.1766863 0.1487130 0.08649266 0.0670
2585
Cumulative Proportion  0.5210822 0.6977685 0.8464815 0.93297415 1.0000
0000
```

These results suggest there may be a second component of interest, even though only one Eigenvalue is greater than 1. A biplot of the measured variables also suggests this:




```
Rotated Loadings for 2 principal components:
    PC1   PC2
I21 0.744 0.413
I22 0.217 0.788
I27 0.158 0.877
I28 0.826
I32 0.711 0.338

               PC1   PC2
SS loadings    1.814 1.675
Proportion Var 0.363 0.335
Cumulative Var 0.363 0.698
```

The two components relate to positive or negative influences on motivation. Items I21, I28 and I32 are assumed to indicate positive motivation and load to PC1 and items I22, I27 and are assumed to be negative motivators and load to PC2.

Moderating Variable 4: “Experience”

Note – with the benefit of hindsight, I can see that the measures chosen for this variable were very poor. These are the only questions not to use a Likert scale. The items did not strongly relate to a single latent construct nor to a standardised measurement scale (initiatives, courses and work environments). The poor scale construction is reflected in the reliability analysis results.

Scale items retained

1. Q33. How many improvement initiatives had you participated in prior to this one?
2. Q34. Have you ever studied quality improvement methods and techniques?

Reliability analysis:

Insufficient items to constitute a Likert scale

Principal Components Analysis

```
Importance of components:
                          Comp.1    Comp.2
Standard deviation     1.1634880 0.8039252
Eigenvalue             1.35       0.64
Proportion of Variance 0.6768522 0.3231478
Cumulative Proportion  0.6768522 1.0000000
> loadings(fit)

Loadings:
    Comp.1 Comp.2
I33  0.707 -0.707
I34  0.707  0.707
```

Comment:

1 significant Eigenvalue, both items load satisfactorily to PC1.

Common Method variance test

Harman single factor test, less than 50% of variance explained by 1 component

```
> fit <- princomp(ImputedPLSPM, cor=TRUE)
> summary(fit)

Importance of components:
                         Comp.1     Comp.2     Comp.3     Comp.4     Comp.5
Standard deviation     3.048260 1.62687775 1.34848452 1.16473831 1.09935132
Proportion of Variance 0.344144 0.09802708 0.06734854 0.05024501 0.04476198
Cumulative Proportion  0.344144 0.44217108 0.50951962 0.55976463 0.60452661
                           Comp.6     Comp.7     Comp.8     Comp.9    Comp.10
Standard deviation     1.04645321 1.02793514 0.95245354 0.89282674 0.85617887
Proportion of Variance 0.04055794 0.03913521 0.03359881 0.02952369 0.02714971
Cumulative Proportion  0.64508454 0.68421975 0.71781856 0.74734225 0.77449196
                          Comp.11    Comp.12    Comp.13    Comp.14    Comp.15
Standard deviation     0.81441376 0.78941277 0.77852272 0.72221129 0.68714882
Proportion of Variance 0.02456555 0.02308046 0.02244806 0.01931812 0.01748791
Cumulative Proportion  0.79905751 0.82213797 0.84458603 0.86390415 0.88139205
                          Comp.16   Comp.17    Comp.18    Comp.19    Comp.20
Standard deviation     0.66145927 0.6313741 0.61596321 0.61255173 0.53914563
Proportion of Variance 0.01620475 0.0147642 0.01405225 0.01389702 0.01076585
Cumulative Proportion  0.89759681 0.9123610 0.92641325 0.94031028 0.95107613
                           Comp.21     Comp.22     Comp.23     Comp.24    Comp.25
Standard deviation     0.515723914 0.500393953 0.474315198 0.415018983 0.38130766
Proportion of Variance 0.009850784 0.009273856 0.008332404 0.006379287 0.00538502
Cumulative Proportion  0.960926912 0.970200768 0.978533172 0.984912459 0.99029748
                          Comp.26     Comp.27
Standard deviation     0.37180386 0.351752709
Proportion of Variance 0.00511993 0.004582591
Cumulative Proportion  0.99541741 1.000000000

> prcomp(ImputedPLSPM, scale = TRUE)
Standard deviations:
 [1] 3.0482598 1.6268777 1.3484845 1.1647383 1.0993513 1.0464532 1.0279351
 [8] 0.9524535 0.8928267 0.8561789 0.8144138 0.7894128 0.7785227 0.7222113
[15] 0.6871488 0.6614593 0.6313741 0.6159632 0.6125517 0.5391456 0.5157239
[22] 0.5003940 0.4743152 0.4150190 0.3813077 0.3718039 0.3517527
```

Appendix 4: PLSPM T Values

PLSPM measurement model	latent variable	Loading	T value
I1. Our initiative identified waste in work processes.	Lean Actions	0.687	8.459
I2. Our initiative identified value-adding activity in work processes.	Lean Actions	0.657	8.176
I3. Our initiative attempted to identify the underlying causes of process problems.	Lean Actions	0.613	8.212
I4. Our initiative observed operational staff in their workplace.	Lean Actions	0.606	6.037
I5. Our initiative used visual tools as part of the solution to support operational processes.	Lean Actions	0.676	10.594
I6. Our initiative attempted to develop safe, reliable and efficient procedures for staff to follow	Lean Actions	0.750	11.533
I7. Our initiative used a Plan-Do-Study-Act framework.	Lean Actions	0.796	8.590
I9. We maintained a focused effort.	Teamwork	0.717	13.478
I11. Team members took responsibility for their actions and behaviours.	Respect for people	0.823	22.358
I12. We avoided personal criticism when reviewing ideas within the team.	Respect for people	0.845	24.021
I14. I enjoyed being part of this team.	Respect for people	0.801	20.627
I15. Team members supported each other.	Respect for people	0.850	29.336
I16. We had clear goals.	Teamwork	0.728	12.959
I17. We had regular meetings.	Teamwork	0.644	9.824
I18. We used facts and data to understand performance.	Teamwork	0.751	13.579
I19. We knew how we were progressing towards our goals.	Teamwork	0.780	15.135
I20. All of our team were encouraged to contribute ideas.	Teamwork	0.726	11.452

PLSPM measurement model	latent variable	Loading	T value
I21. I believed this initiative was an appropriate use of our time and effort.	Motivation 1 (positive indicator)	0.875	26.158
I22. I believed there were better ways to improve the focus area than the solution we chose.	Motivation 2 (negative indicator)	0.852	21.017
I27. Our goals were unrealistic.	Motivation 2 (negative indicator)	0.890	29.188
I28. We felt fully supported by management.	Motivation 1 (positive indicator)	0.673	6.505
I32. I knew why we were attempting this improvement initiative.	Motivation 1 (positive indicator)	0.849	26.302
I38. Our initiative was successful in meeting its stated aim.	Perceived Success	0.906	38.387
I39. Our initiative achieved positive outcomes for patients, staff or the health system.	Perceived Success	0.904	45.126
I41. Please assign a score between 0-100 for your assessment of the success of your initiative.	Perceived Success	0.854	24.297
All P values < 0.00001			

PLSPM structural model	Path Coefficient	T value	P value
Human Resources Capability >> Lean Actions	0.543	6.182	< .00001
Human Resources Capability >> Perceived Success	0.237	2.851	.005294
Lean Actions >>Perceived Success	0.372	4.741	< .00001
Presence of Positive Motivation Indicators >> Lean Actions	- 0.106	0.799	.426184
Absence of Negative Motivation Indicators >> Perceived Success	-0.142	2.261	.025926

Appendix 5: Massey University Human Ethics Committee Low Risk Notification

(Attached file)

MASSEY UNIVERSITY
ALBANY

17 September 2015

William Wilson
33 Taupata Street
Redcliffs
Christchurch 8081

Dear William

Re: Assessing the influence of contextual factors in quality improvement initiatives: an investigation using the model for understanding success in quality (MUSIQ)

Thank you for your Low Risk Notification which was received on 9 September 2015.

Your project has been recorded on the Low Risk Database which is reported in the Annual Report of the Massey University Human Ethics Committees.

You are reminded that staff researchers and supervisors are fully responsible for ensuring that the information in the low risk notification has met the requirements and guidelines for submission of a low risk notification.

The low risk notification for this project is valid for a maximum of three years.

Please notify me if situations subsequently occur which cause you to reconsider your initial ethical analysis that it is safe to proceed without approval by one of the University's Human Ethics Committees.

Please note that travel undertaken by students must be approved by the supervisor and the relevant Pro Vice-Chancellor and be in accordance with the Policy and Procedures for Course-Related Student Travel Overseas. In addition, the supervisor must advise the University's Insurance Officer.

A reminder to include the following statement on all public documents:

> *"This project has been evaluated by peer review and judged to be low risk. Consequently, it has not been reviewed by one of the University's Human Ethics Committees. The researcher(s) named above are responsible for the ethical conduct of this research.*
>
> *If you have any concerns about the conduct of this research that you wish to raise with someone other than the researcher(s), please contact Dr Brian Finch, Director (Research Ethics), telephone 06 356 9099, extn 86015, e-mail humanethics@massey.ac.nz".*

Please note that if a sponsoring organisation, funding authority or a journal in which you wish to publish requires evidence of committee approval (with an approval number), you will have to provide a full application to one of the University's Human Ethics Committees. You should also note that such an approval can only be provided prior to the commencement of the research.

Yours sincerely

Brian T Finch (Dr)
Chair, Human Ethics Chairs' Committee and
Director (Research Ethics)

cc Dr Nihal Jayamaha and Dr Greg Frater
School of Engineering & Advanced Technology

Palmerston North

Professor Don Cleland
Head of School of Engineering & Advanced Technology
Palmerston North