Evaluating Medical Leadership Development Programmes

Oscar Lyons

New College, University of Oxford

Thesis submitted for the degree of

Doctor of Philosophy in Surgical Sciences

February 2021

Abstract

In recent years, effective medical leadership has been recognised across healthcare as being a vital component in the saving of money, staff turnover and patient lives. Increase in the perceived importance of medical leadership has resulted in an explosion of both medical leadership development programmes and investments of public money into these programmes.

These investments into medical leadership development programmes rely on the assumption and conviction that such programmes lead to significant clinical impact. The current lack of effective evaluation of these programmes leaves this an untested assumption. My research has therefore been focused on developing insights into evaluation of medical leadership development programmes.

In Part I of my thesis, I establish a lack of effective evaluation methods for medical leadership development. In Part II, I establish a gap in medical leadership development in medical schools through a needs analysis and a systematic review. In Part III, I shift my focus to qualified doctors and I develop, pilot and critically appraise medical leadership development programme evaluation questionnaires. At the end of Part III, I return to programme stakeholder interviews to more carefully establish their expectations for the medical leadership development programmes and evaluations. In Part IV, to better understand the state of current research and to focus future research, I conduct a second more extensive review of studies evaluating medical leadership development programmes for qualified doctors. Finally, in Part V, I discuss limitations of this research and possible implications for evaluations and future research.

One theme recurs throughout this thesis. If we want to ensure that investment into medical leadership development programmes achieves the ends that commissioners, organisations and politicians intend, we need to get better at questioning our assumptions and evaluating our investments into medical leadership development programmes.

One theme recurs throughout this thesis. If we want to ensure that investment into medical leadership development programmes achieves the ends that commissioners, organisations and politicians intend, we need to get better at questioning our assumptions and evaluating our investments into medical leadership development programmes.

Acknowledgements

Richard and Nick, thank you for taking me on as a doctoral student even though I came to you with almost no research experience. Thank you for your support, care and generosity, for helping to grow my passions for leadership development, and for teaching me how to question my assumptions. I could not have asked for better, more insightful, or more patient supervisors. Martin, thank you for your kindness and encouragement. I will see you as my advisor and mentor no matter which college I'm at. Peter, thank you for your help when I was starting out. Sue, Tom, Trish and Charles, thank you for helping to check my progress and give me direction.

Prof. Hill, thank you for convincing me that a full-time doctorate was a good idea. Dr Buist, thank you for convincing me to apply to Oxford.

The Rhodes Trust, thank you for making it possible for me to step out from full-time practice in NZ into full-time research here in Oxford. Mary, Nadiya, Doyin, Elizabeth, and Charles, thank you for believing in me and for the way you care for all of us.

My colleagues, friends and mentors who contributed to Chapters Four, Five and Ten, working with you has taught me the joy of collaboration in research. I will look forward to bringing your insights, energy and enthusiasm into my research in the years to come. Tatjana, thank you for teaching me how much I have to learn about literature searches! Helen, thank you for your encouragement and for your insights.

Paul Brankin and Nick Hicks, thank you for encouraging me to share my learning with others and for sharing your own ideas with me. Thank you for all the energy you give to your students.

My dear friends Kim, Nick, Bex, Koot, Christy, Steph, Robynne, Harj, Matt, Dre, Jimmy and Sam, thank you for motivating me and helping me figure out how to communicate my research. Thank you for keeping me sane and for distracting me when I needed distractions. Kim, a special thank you for keeping me on track near the end.

João, Deep and Wendy, thank you for your friendship and for running the Emerging Leaders Programme. Peter Sullivan, Claire Pulford, Tony Berendt and Jeremy Noble, thank you for making the ELP possible. To all our ELP graduates, thank you for your patience and your encouragement. Kirstie and Finn, it has been a pleasure working with you, and I am looking forward to seeing what you do with the programme!

Amadea, thank you for your patience, your compassion, and your insights. Thank you for the cups of tea and lunchtime walks, for reminding me to eat when I've been too focused on work and for not minding too much when that work stretches into the early hours. Thank you for sticking with me. Thank you for everything.

Brook, Finn and Jasper, thank you for your constant love and encouragement. Thank you for being my family.

My mum, Brigid, you have invested your life into us. You always told us we could do whatever we set our minds to. You made sure that we had every opportunity, even when it meant considerable hardship for you. You showed us how to overcome challenges, and you helped us believe in ourselves. You are beyond the best supporter, best mum and best person we could ever have hoped to shape our lives.

This thesis is dedicated to you, Mum.

Thesis Contents

PART I: Introduction, Literature Review and Methodology 1

Chapter One - Introduction 2

Introduction 2

Chapter Introduction 4

From personal experience to international focus 4

The increased support for medical leadership internationally 5

Investment in medical leadership development programmes 8

Effectiveness of medical leadership development programmes 9

Need for this research and significance to the field 11

Central research question 11

Questions not covered by this research 12

Thesis structure 12

Chapter Conclusion 18

Chapter Two – Literature Review 19

Chapter Introduction 19

Definitions of leadership 19

Approaches to leadership research 20

A pragmatic definition of leadership 28

Medical leadership 29

Medical leadership development 31

Chapter Conclusion 32

Chapter Three - Methods 34

Chapter Introduction 34

Ontology 35

Epistemology 36
Methodology 37
Critiquing Utilisation-Focused Evaluation 38
Evaluating training programmes: frameworks 40
Evaluating leadership development programmes: methodologies and methods 46
Evaluating medical education: framework and methodologies 53
Methodological conclusions 54
Methodology for this thesis 55
Navigating my personal involvement in the case studies 57
Chapter Conclusion 59

PART II: Leadership Development in Medical Schools 61

Chapter Four – Needs Analysis of Leadership Development in Australian and New Zealand Medical Schools 62

Note on publication of Chapter Four 62
Contribution statement 62
Chapter Introduction 62
Background to the needs analysis 63
Aims 65
Defining learning outcomes for leadership 65
Methods: 66
Discussion 74
Implications for medical schools 77
Chapter Conclusion 78

Chapter Five: Systematic Review of Leadership Development for Medical Students .....Error! Bookmark not defined.

Note on publication of Chapter Five .....**Error! Bookmark not defined.**

Contribution statement .....**Error! Bookmark not defined.**

Chapter Introduction .....**Error! Bookmark not defined.**

Methods.....**Error! Bookmark not defined.**

Results.....**Error! Bookmark not defined.**

Discussion .....**Error! Bookmark not defined.**

Implications for further studies of leadership development in medical schools .....**Error! Bookmark not defined.**

Chapter Conclusion .....**Error! Bookmark not defined.**

PART III: Development of Evaluation Methods for Medical Leadership Development Programmes.....Error! Bookmark not defined.

Chapter Six – Designing a Leadership Development Questionnaire Error! Bookmark not defined.

Chapter Introduction .....**Error! Bookmark not defined.**

The Original Primary Colours Questionnaire.....**Error! Bookmark not defined.**

Critique of the Original Primary Colours Questionnaire....**Error! Bookmark not defined.**

Design of the Revised Primary Colours Questionnaire ....**Error! Bookmark not defined.**

Methods.....**Error! Bookmark not defined.**

Results from Emerging Leaders Programme Cohort One**Error! Bookmark not defined.**

Discussion .....**Error! Bookmark not defined.**

Chapter Conclusion .....**Error! Bookmark not defined.**

Chapter Seven – Retrospective Self-Assessments 81

Chapter Introduction 81

Response-shift bias 82
Retrospective ‘then-test’ 82
Objectives 83
Evaluation methods 84
Analysis 86
Results 86
Additional measures of impact 89
Kirkpatrick Level Two: learning 89
Kirkpatrick Level Three: behavioural changes 89
Kirkpatrick Level Four: organisational results 90
Discussion 91
Chapter Conclusion 95
Chapter Eight – Supplementary Questionnaire Development 98
Chapter Introduction 98
Methods 99
Phase One: Scoping 99
Phase Two: Questionnaire development 100
Phase Three: Questionnaire administration 101
Case studies: 101
Results 103
Free text comments 105
Discussion 109
Chapter Conclusion 113
Chapter Nine: Utilisation-focused evaluation 115
Chapter Introduction 115
Methods 116

Results....120
Discussion....123
Chapter Conclusion....125

PART IV: Systematic Review of Medical Leadership Development Programmes for Qualified Doctors....127

Chapter Ten: Evidence-Based Medical Leadership Development: A Systematic Review....128

Note on publication of Chapter Ten....128
Contribution statement....128
Chapter Introduction....129
Methods....130
Results....136
Discussion....145
Chapter Conclusion....149

PART V: Discussion....151

Chapter Eleven – Discussion, Limitations, Conclusions and Implications for Practice and Research....152

Part I – Introduction, Literature Review and Methodology....152
Part II – Leadership development in medical schools....154
Part III – Development of evaluation methods for medical leadership development programmes....155
Part IV - Systematic Review of Medical Leadership Development Programmes for Qualified Doctors....160
Discussion....161
Limitations....164

Research Implications 166
Practical implications for evaluation of leadership development 166
Areas for further research 167
Personal Reflection 169
References 170

Appendices 209
Appendix One – Supplementary data table from Chapter Four 210
Appendix Two – Supplementary data for Chapter Ten 218
Appendix Three – Original Primary Colours Questionnaire 243
Appendix Four – Revised Primary Colours Questionnaire 250
Appendix Five – Supplementary Questionnaire 253
Appendix Six – Case study descriptions 256

Tables and Figures

Tables

Table 1: Selected approaches to the study of leadership.....21

Table 2: Kirkpatrick's Training Programme Outcome Framework.....41

Table 3: Prominent quantitative methodologies for evaluating leadership. 49

Table 4: Methods for evaluation, compiled and adapted from Hannum (2004).....51

Table 5: Example of mapping and matching results.....72

Table 6: MLCF elements deemed suitable for learning before graduation that are missing from AMC Outcomes for Graduates. 73

Table 7: Key MLCF themes missing from AMC Outcomes for Graduates that are suitable for learning during medical school 74

Table 8: Description of Kirkpatrick's Levels for evaluating educational outcomes. .....**Error! Bookmark not defined.**

Table 9: Summary of studies included in this systematic review**Error! Bookmark not defined.**

Table 10: Comparison between the Original and Revised Primary Colours Questionnaires**Error! Bookmark not defined.**

Table 11: Change in scores across the Revised Primary Colours Questionnaire.....**Error! Bookmark not defined.**

Table 12: Additional methods used to evaluate Emerging Leaders Programme Cohort Two.85

Table 13: Significance tests for Primary Colours Domains, Primary Colours Tasks, and Medical Leadership Competency Framework Domains (ELP2). 87

Table 14: ELP2 Project outcomes. 91

Table 15: Major and minor themes identified in interview with Future Leaders Programme participant 100

Table 16: Demographics of respondents to the Supplementary Questionnaire. 101
Table 17: The NHS Improvement 9C's framework for stakeholder identification (ACT Academy, 2012).....117
Table 18: Results of the systematic text condensation.....121
Table 19: Kirkpatrick's Framework For Evaluation of Training Programmes, with adaptations from Frich et al. (2015).....134
Table 20: Summary of matching between the Medical Leadership Competency Framework and the Australian Medical Council Outcomes for Graduates. 217
Table 21: Summary of included studies for the systematic review in Chapter Ten. 229
Table 22: Study characteristics organised by Medical Education Research Study Quality Indicator (MERSQI) heading. 240
Table 23: Proportion of studies which met Joanna Briggs Institute (JBI) Critical Appraisal Items for Qualitative Studies.....241
Table 24: Case study details 256

Figures

Figure 1: Pendleton and Furnham's Primary Colours of Leadership Model, reproduced from Pendleton and Furnham (2016).....26

Figure 2: Situation of my research within the three methodological literatures explored in Chapter Three. 60

Figure 3: General Needs Analysis Process. 70

Figure 4: PRISMA Flow Diagram.....**Error! Bookmark not defined.**

Figure 5: Change in early to late programme scores in the Original Primary Colours Questionnaire Future Leaders Programme 2016-2017, presented using median, IQR and range.**Error! Bookmark not defined.**

Figure 6: Change in early to late programme scores in the Original Primary Colours Questionnaire, Future Leaders Programme 2016-2017, using mean and standard error. .....**Error! Bookmark not defined.**

Figure 7: Change in early to late programme scores in the Original Primary Colours Questionnaire, Future Leaders Programme 2016-2017, using mean and standard error, participant outcomes only. .....**Error! Bookmark not defined.**

Figure 8: Pre and post-programme scores using the Revised Primary Colours Questionnaire .....**Error! Bookmark not defined.**

Figure 9: Explanation of retrospective pre-test (then-test).....84

Figure 10: An example of a post-course question from the Medical Leadership Competency Framework Self-Assessment incorporating both a post-test ("AFTER") and a then-test ("BEFORE"). 1=strongly disagree, 7=strongly agree.....84

Figure 11: Traditional Pre-, Retrospective Then-, and Post- test self-assessments for the Primary Colours Questionnaire and the Medical Leadership Competency Framework.....88

Figure 12: Impact outside the main programme for the Future Leaders Programme (2016-2017, left) and the Emerging Leaders Programme Cohort One (right).....103
Figure 13: Wellbeing impact of three Emerging Leaders Programme cohorts and the Future Leaders Programme Cohort (2016-2017, bottom right).....104
Figure 14: Pre- and post-programme Brief Resilience Scale scores for Emerging Leaders Programme Cohort Three.....104
Figure 15: PRISMA Diagram.....132
Figure 16: Educational Design Components.....138
Figure 17: Relationship between faculty source and programme outcomes.....139
Figure 18: Educational Methods.....141
Figure 19: Medline (OVID) Search Strategy for systematic review in Chapter Ten, January 2020 242

Preface

Before I started my research, I worked clinically as a doctor. I had also worked for several years as a tutor, lecturer and educator, and I had more extensive extracurricular leadership experience than many of my medical peers. My interests in education and in leadership led me to question why there had been a gap in leadership development during my medical training and led me to start this thesis.

My initial interest was in identifying characteristics of effective leadership development programmes that were important for these programmes to be successful. I reasoned that if these characteristics were clear, leadership development could more effectively be incorporated into undergraduate and postgraduate medical training. I began to investigate evaluation methods I could employ to investigate characteristics of successful programmes.

This initial interest relied on an incorrect assumption, that I would easily be able to identify success in a medical leadership development programme. I realised that because of a lack of established evaluation methods, it would be difficult to establish whether a programme was successful in the first place, let alone to clarify characteristics which supported success. To address this underlying assumption, I adapted my research to focus directly on evaluation of medical leadership development programmes.

Later, as I analysed and critically appraised the results of the questionnaires I had developed, I realised that (despite reviewing literature around medical leadership development and conducting a range of informal interviews) the needs of decision-makers regarding the programmes I was evaluating were still largely unclear and unspecified.

All our evaluation efforts will achieve little with decision-makers if we fail to meet their needs and expectations. Before we can conduct an effective evaluation, we need to ensure that we are clear about the purpose of the evaluation and the needs of key decision-makers, so that we can ensure that our evaluation efforts are best able to inform the decisions they make.

PART I:

Introduction, Literature Review and Methodology

Chapter One - Introduction

Introduction

In recent years, effective medical leadership has been recognised across healthcare as being a vital component in the saving of money, staff turnover and patient lives. There is a growing body of research, expert opinions and political opinions that support this position (Bohmer, 2012; Darzi, 2008; Falcone & Satiani, 2008; Goodall, 2011; Spurgeon et al., 2015; Tasi et al., 2019; West et al., 2015).

Increased perceived importance of medical leadership has led to considerable investments of public money. In the United Kingdom, best estimates indicate that NHS trusts spend up to 29% of their organisational development budget on leadership development (West et al., 2015) or as much as £1 billion per year. This trend is limited neither to the United Kingdom nor healthcare. Beer et al. (2015) estimate that worldwide more than $350 billion (US dollars) is spent on leadership development in American companies.

Such expenditure relies on a key assumption that leadership development programmes lead to significant clinical impact. Lack of effective evaluation of these programmes leaves this an untested assumption. There is currently limited evidence that these investments into leadership development have achieved the impact intended in terms of improved medical leadership or improved quality and cost efficiency in healthcare. There is a paucity of published evaluations of medical leadership development programmes, with a particular shortage of high-quality evaluations.

Two important causes for the lack of high-quality research in medical leadership development are suggested throughout this thesis and throughout the literature: lack of established evaluation methods and lack of clarity regarding intended programme

outcomes (Frich et al., 2015; Geerts et al., 2020; Rosenman et al., 2014; Straus et al., 2013).

The evidence-based approach of modern healthcare provides us with an impetus to organisationally support the evaluation of medical leadership development programmes. In addition to helping us understand the effectiveness of our investments in leadership development, organisationally supporting evaluation will facilitate the development of new knowledge around optimal leadership development. New knowledge regarding optimal medical leadership development could have considerable financial and organisational implications, and could improve the quality of the care we provide.

The focus for this thesis is the development of insights into evaluation methodologies for medical leadership development programmes.

I have organised this thesis into five distinct Parts. In Part I of my thesis, I establish a lack of effective evaluation methods for medical leadership development. In Part II, I establish a gap in medical leadership development in medical schools in Australia and New Zealand through a needs analysis of medical school standards and internationally through a systematic review of medical leadership development programmes in medical schools. In Part III, I shift my focus to qualified doctors and using case studies of medical leadership development programmes based in the NHS I develop, pilot and critically appraise medical leadership development programme evaluation questionnaires. At the end of Part III, I retain my focus on leadership development for doctors in the NHS and I return to programme stakeholder interviews to more carefully establish their expectations for the medical leadership development programmes and evaluations. In Part IV, to better understand the state of current research and to focus future research, I conduct a second more extensive review of international studies evaluating medical leadership development programmes for qualified doctors. Finally, in Part V, I summarise my research and discuss limitations and implications for evaluations and future research.

Chapter Introduction

In Chapter One I introduce my research questions and outline the overall scope and structure of the thesis. I explain how my interest in medical leadership development arose in New Zealand, and how I shifted to a focus on medical leadership development programmes in the United Kingdom and internationally. I briefly and broadly introduce what is currently known about medical leadership development and what is not currently known. I explore the need for my research, and I introduce my research question. Finally, I outline the structure for the rest of the thesis.

From personal experience to international focus

In 2014 I graduated with a medical degree and started work in a small hospital in Gisborne, New Zealand (Tairawhiti District Heath). Having studied medicine for six years, including a one-year full-time “Trainee Intern” apprenticeship, I found myself clinically competent for most aspects of my role as a first-year house officer. From extra-curricular experience, however, I had learned enough about leadership and management to realise I was woefully unprepared for these non-clinical aspects of the first-year house officer role.

After a conversation with Professor Des Gorman, chair of Health Workforce New Zealand, I began to explore literature on medical leadership. I realised that my experience was by no means unique. Indeed, research suggests there is as much need for work introducing medical leadership development into medical schools in the UK (Peake et al., 2018) as there is in New Zealand and Australia (Lyons et al., 2019) (see Chapter Four). Further, the gap in medical leadership development persists beyond medical schools into postgraduate training and senior clinical roles (Frich et al., 2015; Kumar et al., 2020; Lyons et al., 2020; Rosenman et al., 2014; Straus et al., 2013; Sultan et al., 2019).

As articulated later in this chapter, several prominent national UK reports have highlighted the need for improved medical leadership development for practicing doctors. There is a clearer link to patient outcomes from medical leadership development programmes for practicing doctors than from development programmes for medical students (Frich et al., 2015; Onyura et al., 2019; Webb et al., 2014; West et al., 2015). With a view to producing research with increased potential for impact, I therefore adjusted my focus: from medical leadership development for medical schools in New Zealand and Australia (Chapter Four) and medical schools internationally (Chapter Five); through to medical leadership development programmes in the NHS (Chapters Six, Seven, Eight and Nine) and internationally (Chapter Ten).

I focused particularly on medical leadership development in the NHS because of the four countries that have had the strongest focus on medical leadership over the last two decades (UK, United States, Canada and Denmark) (Spurgeon et al., 2011), I had best access to medical leadership development case studies in the UK (see Chapter Three).

The increased support for medical leadership internationally

Only 20 years ago, medical leadership was not often discussed, let alone researched. A literature search conducted by O'Connell and Pascoe in 2004 looking for publications concerning leadership development in medical schools found only 15 articles with any degree of relevance (O'Connell & Pascoe, 2004). In 2014, with more stringent inclusion criteria, Webb et al. found 45 articles (Webb et al., 2014). Since Webb, other systematic reviews have been conducted and published regarding medical leadership development programmes in academic centres (Straus et al., 2013), in specialist trainees (Onyura et al., 2019; Sultan et al., 2019), in physicians (Frich et al., 2015) and in medical schools (Lyons et al., 2018). These reviews have each found a small but growing body of literature concerning medical leadership development programmes.

Medical leadership in the NHS

The United Kingdom has had a prominent focus on improving medical leadership over the past two decades, and therefore has been a useful setting for this research. The UK increase in prominence of medical leadership as a concept and as a means to improve quality and affordability in the NHS (National Health Service) has occurred partly in response to national reports and inquiries. As the focus of the thesis is not on the history of leadership development in the NHS, I will briefly cover some of the more significant of these national reports without delving too deeply into the complex political history of medical leadership in the NHS.

The NHS Plan (2000)

The NHS Plan in 2000 was clear about the need for medical leadership, declaring "We need clinical and managerial leaders throughout the health service." The 2000 NHS Plan went on to comment that medical leadership development has been disordered: "Leadership development in the NHS has always been ad hoc and incoherent with too few clinicians in leadership roles and too little opportunity for board members to develop leadership skills. That will now change" (Department of Health, 2000, p. 87). Despite this strong rhetoric, change has been slow to occur.

The Darzi Report (2008)

Perhaps the loudest support for medical leadership in the NHS came eight years after the NHS Plan in the form of the Darzi Report (Darzi, 2008). Lord Darzi advocated that medical leadership had been the "neglected element of the reforms of recent years" (Darzi, 2008, p. 66). His report into the state and the future of the health system concluded that "Quality is improved by empowered patients and empowered professionals. There must be a stronger role for clinical leadership and management throughout the NHS" (Darzi, 2008, p. 48). Calling for increased quality and consistency of leadership development programmes for clinicians, Darzi asserted that "It is unrealistic

to expect NHS staff to take on leadership without action to make it integral to training and development" (Darzi, 2008, p. 72). He acknowledged that there were already local trust investments in such leadership development programmes, while highlighting that these programmes, despite the call to action in the NHS Plan, tend still to be "variable in their scope and standard" (Darzi, 2008, p. 66).

A key development in the UK following the Darzi report was the national programme "Enhancing Engagement in Medical Leadership" established jointly by the Institute for Innovation and Improvement and the Academy of Medical Royal Colleges (NHS Institute for Innovation and Improvement, 2010). This project developed the Medical Leadership Competency Framework (MLCF) as a description of the leadership content needed by doctors to take on leadership roles, the Medical Engagement Scale (to assess the level and impact of degrees of medical engagement within the medical workforce), and led the drive to establish the new Faculty of Medical Leadership and Management (FMLM) (Spurgeon et al., 2011).

The Faculty of Medical Leadership and Management

The Faculty of Medical Leadership and Management (FMLM) was established in 2011 to improve the cohesiveness and prominence of medical leadership development. The FMLM was established by the presidents of all the Royal College of Physicians, the Royal College of Surgeons, and the Chair of the Royal College of General Practitioners and was formally endorsed by the Academy of Royal Colleges in January 2011. The FMLM maintains a vision to "see and inspire excellence in medical leadership and drive continuous improvement in health and healthcare in the United Kingdom" (Faculty of Medical Leadership and Management, 2017).

The Francis Report (2013) and the Nuffield Trust Review (2016)

In more recent years, the 2013 Francis Report of the Mid Staffordshire NHS Foundation Trust Public Inquiry continued the call for the development of medical

leadership (Francis, 2013). The Nuffield Trust's independent review in 2016 found that significant problems remain in terms of relationships between doctors and managers, and in terms of support for managers in terms of their environment and resources. "If improving working relationships between doctors and managers is an important strand of building a strong service for the future (as we believe it is), then NHS managers – both medical and non-medical – need to be valued, equipped with the necessary skills and resources and provided with a context that enables them and the clinical services that they manage, to flourish. Evidence from the medical and non-medical managers in this study suggests that there is a long way to go" (Powell & Davies, 2016).

There has been much discussion of improving medical leadership development. As I will describe next, this discussion has not necessarily achieved the changes intended or expected.

Investment in medical leadership development programmes

Investment and number of programmes internationally

There are now substantial investments into medical leadership development. Best estimates suggest that up to 29% of organisational development budgets in NHS trusts is spent on leadership development, which would amount to more than £1 billion pounds sterling per year in the NHS alone (West et al., 2015). To put this figure into perspective, £1 billion per year could fund 26,000 additional nurses.

Investment in leadership programmes is neither specifically an NHS trend nor confined to healthcare. Beer et al. (2015) estimate that across companies worldwide the annual spend on leadership development exceeds $350 billion US. As the perceived importance of leadership development has increased, so has the number of leadership development programmes and the investments they have attracted (Beer et al., 2015). There is now an imperative that every medical school in the United Kingdom ensures

their students can demonstrate the competencies outlined in the Medical Leadership Competency Framework (MLCF) before they graduate, and professional standards for doctors also now include leadership at all levels (Clark & Armit, 2008; McKimm & Swanwick, 2011).

Despite the increased investments into medical leadership development programmes, there appears to be limited evidence that these investments into leadership development have achieved the impact intended in terms of improved quality and cost efficiency of healthcare.

Effectiveness of medical leadership development programmes

Lack of evidence

Despite modern healthcare priding itself on being evidence-based, little is known about the effectiveness of leadership development interventions in healthcare (Clark & Armit, 2008; Leslie et al., 2005; Straus et al., 2013; West et al., 2015). This lack of evidence regarding the effectiveness of leadership development extends outside healthcare (for example, Beer et al., 2015; Day & Sin, 2011). West et al., in their 2015 review of the evidence base for medical leadership development, lament that “Overall, the evidence for the effectiveness of specific leadership development programmes within the NHS is highly variable and little robust evidence has been accumulated, despite the vast sums spent” (West et al., 2015).

Lack of evaluation

This lack of evidence in part stems from a lack of medical leadership development research, and in part from a lack of effective evaluation methods (Frich et al., 2015; Geerts et al., 2020). While many bodies have attempted to produce medical leadership frameworks (for example, the Academy of Medical Royal Colleges, NHS Leadership Academy, Faculty of Medical Leadership and Management, LEADS Canada), there is

little evidence to support the use of any particular framework, and most of these frameworks are based on unclear theoretical and psychometric foundations (West et al., 2015). Leadership interventions continue to be based primarily on popular literature and the opinions of educational providers, with “perpetual discussions but very little research” (Powell & Yalcin, 2010). Evaluations and research have relied primarily on commentaries, self-reports and cross-sectional surveys, which makes it difficult to extrapolate practical implications from individual studies (Clark & Armit, 2008; Frich & Spehar, 2018; Leslie et al., 2005; Lyons et al., 2018; West et al., 2015). The quality of evaluations and research will be explored in greater detail in Chapter Two.

Consequences of lack of evaluation

Evaluation of any programme is essential to curriculum development (Patton, 2008; Thomas et al., 2016; Watkins et al., 2011). Lack of evaluation has consequences for the effectiveness of leadership development interventions. Leslie (2005) has described current leadership development as “sporadic and rudimentary”; Clark and Armit (2008) as “at best scant and ad hoc” (in a stark echo of the NHS Plan’s declaration); and Satiani et al. (2014) wrote that “what passes as leadership development in some hospitals and medical schools is a hodgepodge of classes and lectures lacking coherence, logical progression, comprehensiveness, and relevance”.

In a broader sense, while published studies have tended to show positive effects, there are certain individual medical leadership development programmes that seem to have had little or no effect (Frich et al., 2015; Lyons et al., 2018; West et al., 2015). Poor reporting of content, design and evaluation strategies (Frich et al., 2015; Straus et al., 2013), renders it difficult for readers and researchers to determine why some studies have shown poor outcomes. It is difficult to ascertain whether reduced impact might result from quality of individual programmes, differences in evaluation methodologies, or other factors entirely.

Need for this research and significance to the field

Lack of research regarding medical leadership development

There appears to be consensus in the medical field that there is need for more optimal medical leadership development. In order to optimise and propagate best practices in medical leadership development, we clearly need good evidence as to what best practice is. Good evidence regarding best practice requires appropriate, reliable, valid and feasible approaches to research and evaluation.

Uncertainty regarding optimal evaluation methods

Broad uncertainty remains as to how best to evaluate leadership development programmes (Frich et al., 2015; Leslie et al., 2005; Straus et al., 2013; West et al., 2015). Lack of accepted and appropriate evaluation methodologies is likely to be a factor which hinders the efforts and motivation of programme providers to evaluate their programmes and to meaningfully contribute to the body of leadership development data.

Central research question

At the outset of this research, my primary aim was to explore factors which influenced and predicted positive outcomes from medical leadership development programmes.

As my understanding of medical leadership development programmes progressed, it became clear that my initial aim was too far downstream. It relied on an assumption that there were established effective methods for evaluating medical leadership development programmes. It became clear early in my research that such methods have not yet been established within the literature. Addressing the literature gap concerning best practice in medical leadership development programmes therefore relies on an upstream step: improvement of leadership development evaluation methodologies.

My central research question therefore became:

"How can evaluation of medical leadership development programmes be improved?"

Questions not covered by this research

As highlighted throughout this thesis, our understanding of medical leadership development is still nascent. Questions remain, including: what the ideal design of a leadership development programme is; at what stage leadership should best be taught; what content is most important; what the optimal makeup of programme faculty is. These topics are important but go beyond the scope of this thesis.

My aim is to contribute insights regarding effective evaluation methodologies into the discourse on medical leadership development. More effective evaluation methodologies may in the future facilitate research into those broader questions.

Thesis structure

PART I: Introduction, Literature Review and Methodology

In Part One I introduce the narrative of this thesis. I review literature around leadership, leadership development and medical leadership development, I explore the approaches which have been taken to researching these areas, and I explain my methodological approach.

Chapter One - Introduction

In Chapter One I provide context for the development of the research question, introduce the research questions and outline the significance of my research to the emerging medical leadership development field.

Chapter Two – Literature Review

In Chapter Two, I provide further context for my research, outline what has been accomplished thus far in medical leadership development research, and situate my new research within the established medical leadership development literature. I explore some of the main definitions of leadership and medical leadership which have developed over the last century. I then outline a pragmatic working definition of leadership and of medical leadership for my research.

Chapter Three - Methodology

In Chapter Three, I introduce the overall approach and research philosophy which underpins my research as presented in this thesis. I then explore existing approaches to evaluating training programmes, leadership development programmes, and medical education. I critically evaluate the assumptions and the appropriateness of these existing methodologies for my utilisation-focused evaluation research in medical leadership development programmes. I then describe the mixed methods that I have chosen for my research and I explain the case studies chosen for my research.

PART II: Leadership Development in Medical Schools

In Part II, I focus on leadership development for medical students and I establish a gap in leadership development in medical training in Australia and New Zealand and internationally. Chapter Four focuses on Australia and New Zealand, which was where I trained. Chapter Five then moves to look at international medical leadership development in medical schools, before Part III refocused on case studies in the United Kingdom.

Chapter Four – Needs analysis of medical leadership in medical training

In Chapter Four, I present the results of a systematic needs analysis of the Australian and New Zealand Medical Curriculum Standards, in order to assess the degree to which incorporation of further leadership development into this curriculum is needed. The results of this chapter show a need for increased attention to leadership development in medical schools.

Chapter Five – A Systematic Review of Leadership Development for Medical Students

In Chapter Five, I present the results of a systematic review of the literature regarding leadership development for medical students. The purpose of this chapter is to evaluate the current state of the literature and to situate my early research within the established body of literature. This chapter shows that internationally there have been few evaluations of medical leadership development programmes in medical schools, and highlights marked heterogeneity in the research that has been published.

PART III: Development of Evaluation Methods for Medical Leadership Development Programmes

After establishing a gap in medical leadership development for medical students, I found that current literature suggests that this gap persists into the clinical practice of qualified doctors. In Part III I move from leadership development for medical students into development of evaluation methods for medical leadership development programmes for qualified doctors. In Part III I use case studies of five cohorts across three leadership development programmes based in Oxford University Hospitals to develop, pilot, appraise and improve evaluation methods.

Chapter Six – Development of a Leadership Competency Self-Assessment

In Chapter Six, I introduce the Original Primary Colours Questionnaire developed to evaluate the Future Leaders Programme at Oxford University Hospitals Foundation Trust. Through critical analysis, I explore the reasons for my redesign of this questionnaire into the Revised Primary Colours Questionnaire. I then present the results from application of the Revised Primary Colours Questionnaire into the first Emerging Leaders Programme cohort, alongside qualitative data collected through questionnaires, interviews and documents. I critically appraise these results with a particular emphasis on elucidating missing components, unexpected findings and limitations of this approach.

Chapter Seven - Retrospective pre-post self-assessments

In Chapter Seven, I introduce the technique of Retrospective Pre-Post Self-Assessment as a method for reducing response-shift bias. I briefly explore the historical use of Retrospective Self-Assessment in other settings, and I explain my rationale for applying this technique to medical leadership development programme evaluations. I then present the results of Retrospective Self-Assessment from the evaluation of the second Emerging Leaders Programme cohort. I critique the effectiveness of this approach with reference to other data I collected by interviews and documents, as well as through co-administration of a parallel questionnaire adapted from the Medical Leadership Competency Framework Self-Assessment Tool. I again place emphasis on a need for elucidating missing components of the evaluation, exploring unexpected findings, and addressing limitations.

Chapter Eight - Development of the supplementary questionnaire

In Chapter Eight, I introduce the background to the supplementary questionnaire, which assesses participant application of learning outside the programme, involvement

in other projects and perceived effect on measures of confidence, motivation, job satisfaction and resilience. This Supplementary Questionnaire was developed in response to participant interviews from the Future Leaders Programme. Consultation with programme participants drew attention to significant outcomes that were not being captured by the Primary Colours Questionnaires. These related to extra-curricular involvement in quality improvement and service redesign, as well as to the personal impact of the programmes for participants. In Chapter Eight, I explore the development of this supplementary questionnaire, present results from its application to four case studies, and align these results with data from free text comments, interviews and focus groups.

Chapter Nine - Utilisation-focused evaluation

Tying together Chapters Six, Seven and Eight, in Chapter Nine I summarise the findings from the five case studies and I reconsider the utility of these evaluation methodologies. In order to establish how appropriate and complete these findings were for utilisation by commissioners, funders, providers, faculty and participants, I investigate the perceived and reported expectations for evaluation from interviews with a range of stakeholders in medical leadership development programmes. These interviews focused on the purpose of involvement of interviewees in respective medical leadership development programmes; their intended outcomes from these programmes and their perception of what success would look like in medical leadership development programmes. The interviews presented in Chapter Nine contextualise the findings from preceding chapters and highlight potential areas for improvement of evaluations of future medical leadership development programmes.

PART IV: Systematic Review of Medical Leadership Development Programmes for Qualified Doctors

Part IV returns to a second systematic review of leadership development, with a focus on qualified doctors and a renewed emphasis on critical appraisal of programmes and of evaluation methods.

Chapter Ten – Evidence-Based Medical Leadership Development: A Systematic Review

In Chapter Ten, I present the results of a second systematic review which focuses on leadership development for qualified doctors. This review gives a more complete context for the leadership development case studies presented in Chapters Six, Seven and Eight, and draws attention to findings which extend across the body of literature.

PART V: Discussion

Chapter Eleven - Discussion, Limitations and Implications for Practice and Research

In Chapter Eleven I briefly summarise the conclusions I have drawn from my research across the chapters in Parts I-IV. I then discuss my findings, comparing and contrasting with the established literature. I highlight new and unexpected findings, and critically appraise the strength and limitations of my research. Finally, I consider implications for future evaluation of medical leadership development programmes, as well implications for future research into evaluation of medical leadership development programmes.

Appendices

The appendices contain supplementary material from Chapter Four and Chapter Ten, copies of the original and the revised Primary Colours Questionnaires and of the

Supplementary Questionnaire, as well as details of the five case studies which Part III of this thesis focuses on.

Chapter Conclusion

The need for research into medical leadership development programme evaluation

In Chapter One I have introduced the research question for my thesis and explained the context from which this question developed. I have also laid out the structure of this thesis and have emphasised how it falls into five Parts. In Chapter Two, I will further explore the literature surrounding evaluation of medical leadership development, and I will outline the gap in the literature that my research addresses.

In Chapter Two, I will further explore the literature surrounding evaluation of medical leadership development, and I will outline the gap in the literature that my research addresses.

Chapter Two – Literature Review

Chapter Introduction

In Chapter One I explained how and why my interest in medical leadership evolved into the specific research questions for this thesis. I described the rise of interest in medical leadership development programmes and introduced the gap in research concerning evaluation of these programmes which this thesis will explore.

Alongside a lack of clarity regarding effective evaluation methods, there is a lack of clarity regarding what “leadership” means. In Chapter Two I briefly explore the definitions and history of leadership and medical leadership.

The purpose of Chapter Two is to provide context for the research in terms of leadership generally and medical leadership specifically and to introduce gaps in the established literature.

Definitions of leadership

Theories and definitions of leadership have proliferated and evolved over the years (Parry & Bryman, 2013). Stogdill famously lamented in 1974 “there are almost as many different definitions of leadership as there are persons who have attempted to define the concept” (Stogdill, 1974). Adding to this confusion, there are a multitude of lay definitions of leadership which are implicitly or explicitly held. A unifying definition of leadership remains elusive.

Some examples of definitions follow:

"The process (act) of influencing the activities of an organised group in its efforts towards goal setting and achievement"

(Stogdill, 1950).

"The art of motivating a group of people to achieve a common goal"

(The King's Fund, 2011, p. 12).

"The projection of personality and purpose onto people and situations in order to prevail in the most demanding circumstances"

(Ministry of Defence, 2010).

"To create the conditions for people to thrive, individually and collectively, and achieve significant goals"

(Pendleton & Furnham, 2016).

Attempting to provide a unifying definition is outside the scope of this thesis. I will, rather, present some of the prominent theories and definitions which have been used and introduce a pragmatic working definition for this thesis.

Approaches to leadership research

As study of leadership has become more popular, various approaches have been taken to researching and understanding leadership. Parry and Bryman (2013) explain that each new approach has changed the emphasis of study, rather than subsuming previous approaches. This has resulted in a continued proliferation of approaches. Parry and Bryman describe five distinct approaches to leadership theory, which have been supported by Northouse (2013) and further developed by Dopson et al. (2016). Across these five approaches, there has been a tension between focusing on individuals and focusing on context. To illustrate these shifts and the key added value of each approach,

I have reorganised Parry and Bryman's approaches into five categories, including more recent research approaches where possible, as shown in Table 1 and explored below.

Approach	Prominent Years	Individual/Context focus
Trait	1940s	Individual
Task and Relationships	1940s-1980s	Context
Managers of Meaning	1980s-present	Individual
Distributed Leadership	2000s-present	Context
Fragmentation of approaches	Present	Mixed

Table 1: Selected approaches to the study of leadership
Adapted from categories proposed by Parry and Bryman (2013)**. This table illustrates the development of new approaches and shows the tendency for approaches to overlap. The Individual/Context focus column simplifies these approaches (as many have looked at both individuals and the context that they exist within) but has been included to show the tension between approaches that focus on individuals or context.**

Traits

The Trait approach, which focused on traits of individual leaders, remained prominent up until the late 1940's (Day, 2014; Parry & Bryman, 2013). Perhaps the most important exposition of the trait approach was Stogdill's 1948 review (Stogdill, 1948). In this review, Stogdill synthesised material from more than 120 primary sources to quantify the support for each of 29 traits. Most strikingly, there was extreme and often conflicting variation between individual study findings. Study quality was also low and at times "the methods used in the investigation appeared to have little relationship to the problem as stated" (Stogdill, 1948). Consistency was found in a small number of traits (intelligence, scholarship, dependability, activity and social participation, and socio-economic status) (Stogdill, 1948), though these traits raise definitional questions of their own.

In an early hint of upcoming change in approaches to studying leadership, Stogdill noted that "the qualities, characteristics, and skills required in a leader are determined to a large extent by the demands of the situation in which he is to function as a leader" (Stogdill, 1948). While the studies Stogdill explored primarily concerned whether an individual's traits would affect whether that individual was regarded as a leader in different contexts, the same studies also noted that different traits could be observed in

the same individual in different situations. The conclusion that context can affect the manifestation of leadership traits led to the development of Style and Contingency approaches to leadership.

Relationships and Tasks: The Style and Contingency approaches

In the late 1940's, the style approach began to emerge from Stogdill's Ohio State University. This new approach focused less on inherent personal characteristics of leaders and more on the behaviour of leaders. The new and distinguishing feature was separation of leadership into two axes of behaviour: task-orientation and relationship-orientation (Northouse, 2013; Parry & Bryman, 2013). The aim of the style approach was to understand and explain "how leaders combine these two kinds of behaviours to influence subordinates in their efforts to reach a goal" (Northouse, 2013, p. 75).

The style approach signified a shift in leadership development as well as leadership theory. Focusing on behaviours rather than on "traits" allowed for a shift in perception. Because this approach recognised that leadership behaviours could be changed, organisational leadership development efforts began to shift from leadership talent identification and selection towards training and development (Parry & Bryman, 2013).

Where the Ohio State group focused on a leader's ability to change their style to suit a context, Fred Fiedler's Contingency Model, developed in the 1960s, focused on a leader's suitability for a context (Fiedler & Chemers, 1974). Fiedler used similar terminology to the style approach, but rather than using scales of relationship and task orientation to describe means of behaving, he considered them fixed measures that defined an individual. This could be viewed as a regression to an individual-focused approach as in the trait approach. Fiedler's nuance was his argument that the effectiveness and success of an individual leader is contingent on how favourable the situation is to the leader (cf. style approach).

Managers of Meaning: The New Leadership approaches

Whereas previous approaches had focused on individual leaders, approaches that emerged in the 1980's began to focus on interactions between leaders and followers, and specifically on how leaders convey vision and meaning to their followers (Northouse, 2013; Parry & Bryman, 2013). These "new leadership" approaches were derived mainly from studies of influential and successful senior leaders and included charismatic leadership and transformational leadership. These still-popular approaches exemplify effective leadership as providing direction by transforming followers; changing and activating their personal values and inspiring internal motivation from self-interest to collective interest (Northouse, 2013; Parry & Bryman, 2013).

James MacGregor Burns' 1978 study of political leadership first introduced a dichotomy of transactional and transformational leadership (Burns, 1978). This dichotomy was further developed by Bennis and Nanus (1986) and Bass (1985), and contributed to the popular separation of leaders (transformational leaders) and managers (transactional leaders). I will briefly address leadership and management in a later section.

Bass (1985) viewed transformational and transactional leadership as separate and concurrent scales rather than opposite ends of the same scale. Transactional leadership broadly focused on the use of external motivations such as payment or rules to influence followers, whereas transformational leadership broadly focused on inspiring internal motivation in followers through introduction of new perspectives and through alignment of vision. Research into transformation and transactional leadership approaches (Den Hartog et al., 1997; Waldman et al., 1990; Yammarino et al., 1993) suggested that transactional leadership was by itself necessary but not sufficient for optimal performance. "Transactional leadership provides the base for effective leadership and performance at expected standards, while transformational leadership provides the add-

on for superlative leadership and performance beyond expectations"(Yammarino, 1993). Charismatic leadership approaches aligned closely with transformational leadership approaches, in that leaders were considered to transform the values, aspirations and needs of their followers from a self-focus to a collective focus (Dopson et al., 2016).

Both charismatic and transformational leadership approaches have been criticised for lacking attention to ethics and have been condemned by Van Knippenberg and Sitken (van Knippenberg & Sitkin, 2013), who question concept definitions, measurement tools, results, conclusions and applicability. Given the breadth of concerns regarding charismatic and transformational leadership, Van Knippenberg and Sitken recommend that charismatic and transformational approaches be abandoned, and more clearly defined and distinct aspects of leadership be studied instead.

Distributed Leadership

Distributed leadership approaches represent a further step away from a focus on individual leaders towards a perspective that considers leadership to exist in the space between leaders, followers and context. These approaches "project an image of leadership as much more diffuse and dispersed within organisations than would be evident from the tendency for leadership to be viewed as the preserve of very few leaders as in [previous approaches]" (Parry & Bryman, 2013). In distributed leadership approaches leadership is viewed as a "concertive action" with varieties of expertise distributed across many individuals (Bennett et al., 2003, p. 7). Two particular strands emerged: group leadership and followership.

Group leadership

Studying leadership within e-commerce, Brown and Gioia described a system of group leadership, whereby leadership is not the sole province of leaders, nor described by a set of characteristics of such persons. Rather, they explored leadership as a

property of a social system which emerges as both leaders and followers share in the process of moving towards goals and enacting leadership (Brown & Gioia, 2002).

Pendleton and Furnham's research with the Edgecumbe Group expanded on this concept of group leadership and provided some empirical evidence to support the approach (albeit by study of senior leaders rather than systems) (Pendleton & Furnham, 2016). They conducted executive assessments at three major organisations with 100 executives. They measured the capabilities of these executives against their 'primary colours' model, which separates leadership into three domains (strategy, interpersonal and operational) with seven overlapping components and an additional "coping with pressure" component.

Pendleton and Furnham's assessments found that no individuals were outstanding in all components, and those who were outstanding in one component were frequently measured to have significant shortfalls in another component. The shortfall, they argued, is made up by others in a leadership team. This argument supports Rath and Conchie's assertion that "although individuals need not be well-rounded, teams should be" (Rath & Conchie, 2008).

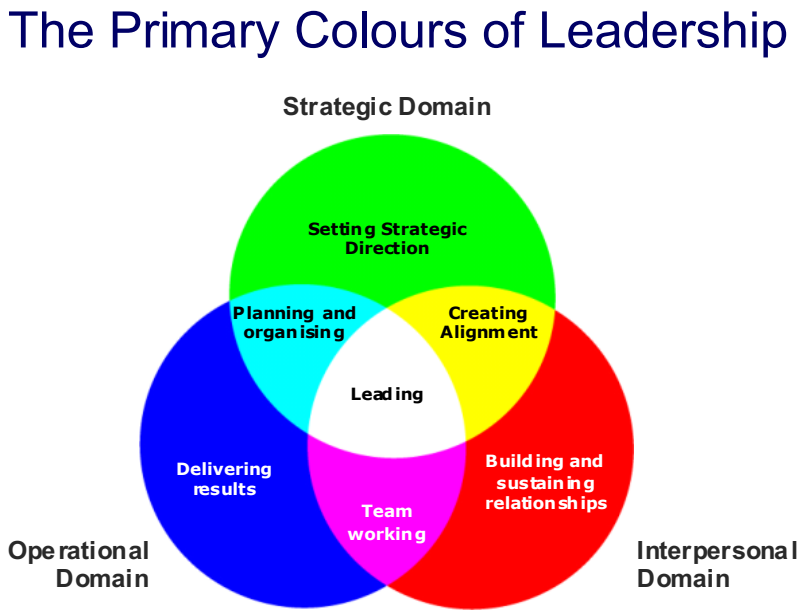


Figure 1: Pendleton and Furnham's Primary Colours of Leadership Model, reproduced from Pendleton and Furnham (2016)

Followership

The followership approach has roots in the works of Mary Parker Follett (1868-1933). Follett advocated that ideas and means of improvement could come from followers as well as leaders when "leader and followers are both following the invisible leader – the common purpose" (Follett, 1949, p. 55). There has been a recent increase of interest in followership after Kelley's Harvard Business Review article "In praise of followers" (2011). Mintzberg, for example, asserts that too much focus is placed on leadership, neglecting the contribution of followers (Mintzberg, 2004). Grint (1997) quotes Tolstoy's War and Peace analogy of leadership as a bow wave to illustrate a similar principle: a bow wave precedes a ship but it does not move the ship, rather it is created by the ship.

In 2014, Mary Uhl-Bien and colleagues moved to define followership clearly and situate it with respect to previous approaches to leadership (Uhl-Bien et al., 2014). Uhl-Bien separates followership research into two approaches: role-based approaches which investigate the influence of followers on leadership behaviours; and constructionist approaches which look at how followers co-create leaders through social processes. Both followership approaches assert that individuals within an organisation who are

normally designated as followers have a vital role in informing and steering leaders and the organisation.

Followership approaches have attracted interest from those studying and promoting leadership in healthcare. There has been an increasing effort to empower healthcare team members traditionally perceived to be “just” followers within established hierarchies to speak up and influence team and leader behaviours, particularly in crises (Bromiley, 2015; McKimm & Vogan, 2020; Morrow et al., 2016).

Fragmentation of approaches

In the last twenty years, there has been an upsurge in approaches which attempt to take a broad view of leadership, acknowledging that previous quantitative and rational approaches have failed to account for the complexity of leadership (Avolio, Walumbwa, et al., 2009). There are a wide range of new approaches to studying and understanding leadership which have recently proliferated, including (to name just a few), pragmatic leadership (Mumford & Van Doorn, 2001), spiritual leadership (Fry, 2003), art of leadership (Grint, 2000), authentic leadership (Walumbwa et al., 2008) and complexity leadership (Uhl-Bien et al., 2007).

Avolio and colleagues point out that “never before has so much attention been paid to leadership” (Avolio, Walumbwa, et al., 2009). It is perhaps not surprising that the proliferation of leadership research, combined with the lack of agreed best approaches to researching and understanding leadership in general (Ardichvili et al., 2016) or in healthcare specifically (West et al., 2015), has resulted in fragmentation of approaches to leadership.

It remains to be seen which approaches will withstand the test of time. This presents a dilemma for contemporary research regarding leadership: it is difficult to clearly define

and evaluate successful leadership development when there remains no clear common definition of leadership or common approach to studying leadership.

"Leadership" and "Management"

As with leadership, management does not have a commonly agreed definition (Northouse, 2013; Parry & Bryman, 2013; Yukl, 2013). There has been debate as to what constitutes leadership and what constitutes management since the introduction of a dichotomy between transactional and transformational leadership (see the 'Managers of Meaning' section above). From a research perspective, perhaps the most commonly held position is that leadership and management are distinct roles that need not necessarily reside in more than one person (Kotter, 1990; Mintzberg, 1980; Yukl, 2013). There are however indications that excellence of a single individual in both leadership and management roles is not common, and this has perhaps strengthened the ongoing debate (Kotter, 1990; Pendleton & Furnham, 2016). It is clear, however, that many functions and competencies seem to be shared by leader and manager roles (for example, M. Young & Dulewicz, 2008). For the purpose of this thesis, it is not vital to establish the distinction between leadership and management. As such, I will consider the term "leadership" to broadly include the responsibilities of management, as in Pendleton and Furnham (Pendleton & Furnham, 2016).

A pragmatic definition of leadership

Given lack of an agreed approach to defining leadership, Yukl contends that "leadership research should be designed to provide information relevant to a wide range of definitions" (Yukl, 2013, p. 7). It seems appropriate therefore to adopt a broad pragmatic definition that fits the basic conditions for leadership and for leadership research. Stogdill perhaps has offered the broadest articulation of the basic conditions for leadership. He states these conditions to be:

"1. A group (of two or more persons).

2. A common task (or goal-oriented activities).

3. Differentiation of responsibility (some of the members have different duties)"

(Stogdill, 1950).

In this thesis, with a view to producing research that is widely relevant and fits Stogdill's basic conditions, I will take a broad approach to leadership, which aligns with Blake's definition of leadership: *"*achieving results with and through others*"* (Blake & Mouton, 1985, p. 198).

Medical leadership

Given the lack of an established agreed definition of leadership generally, it does not come as a surprise that there is also no agreed definition of medical leadership specifically.

In the medical leadership literature, two broad streams of research can be identified that have different approaches to defining medical leadership (Berghout et al., 2017). The first considers medical leadership to mean doctors taking on formal leadership roles such as clinical lead, clinical director and medical director. (Andersen, 2006; Goodall, 2011; Sarto & Veronesi, 2016; Savage et al., 2020). The second stream considers medical leadership to be an intrinsic component of doctors' daily work (Edmonstone, 2009; Frich et al., 2015; Noordegraaf et al., 2015). The boundary between these approaches to medical leadership is usually implicit and often unclear (Berghout et al., 2017). Each approach presents different challenges to research, therefore there is a value and need for both. In this thesis, instead of separating these two streams, I will treat medical leadership as relevant in a continuum from the daily work of doctors through to formal leadership positions.

Confusingly, it is common in the research to conflate clinical leadership with medical leadership. There are however commonly accepted differences between clinical and medical leadership. Clinical leadership usually refers to leadership roles (formal or informal) for any healthcare professional, including doctors, nurses, physiotherapists, occupational therapists, midwives, podiatrists (NHS Improvement, 2019). Medical leadership and physician leadership refer specifically to leadership roles (formal or informal) for those with medical qualifications (regardless of whether they currently practice clinically) (Goodall, 2011; NHS Institute for Innovation and Improvement and Academy of Medical Royal Colleges, 2009).

For the purpose of this thesis I will treat medical leadership as a subset of clinical leadership. This is the position most commonly held in the literature (Spurgeon et al., 2015; West et al., 2015).

Strengthening focus on medical leadership

Calls for improved quality of patient care and increased capacity of healthcare systems to implement and adopt innovations have brought with them increased focus on engaging doctors in change management and in leadership (Bohmer, 2012; Darzi, 2008; Fisher et al., 2009; Goodall, 2011). There are strong expert opinions and political opinions that effective medical leadership is important for healthcare outcomes (Bohmer, 2012; Darzi, 2008). This position is also supported by a growing body of research literature (Geerts et al., 2020; Goodall, 2011; Goodall & Stoller, 2017; Stoller, 2009), despite persistent challenges from the lack of agreed understanding as to what leadership means in this context.

It has been suggested that the shift towards increased formal medical leadership has been curtailed by the reluctance of doctors to take on leadership and management responsibilities (Aggarwal & Swanwick, 2015; Ireri et al., 2011; Spehar et al., 2012). Clinicians have long considered management and leadership to be the “dark side”

(Spurgeon et al., 2011), where the focus on budgets and administration is at the expense of a focus on patients' quality of care. Supporting this position, Spehar found that one of the main reasons for doctors moving into 'clinician-manager' roles (which he defined as formal leadership roles held by clinicians) was in order to protect their colleagues from non-medical managers who might otherwise step into those roles and maliciously or carelessly inflict harm on their colleagues and the care of their patients (Spehar et al., 2012).

There has nonetheless been an increasing support for medical leadership (Berghout et al., 2017; Ireri et al., 2011; Spehar et al., 2012). Internationally professional bodies have been set up to support and regulate medical leadership, including for example the Faculty of Medical Leadership and Management (FMLM), the Royal Australasian College of Medical Administrators (RACMA), the American Physician Leadership Institute. Medical leadership has been adopted as a core competency in medical councils including the United Kingdom General Medical Council (General Medical Council, 2018), the Canadian Medical Association (Frank JR, Snell L, Sherbino J, 2015), the New Zealand Medical Council and Australian Medical Council (Australian Medical Council, 2012), and a range of other national medical councils and regulators.

Rather than defining what medical leadership means, medical councils have sought to highlight competencies and capabilities of effective medical leadership.

Medical leadership development

While there seems to be agreement that effective medical leadership has a positive effect on healthcare outcomes, it is less clear that medical leadership development programmes have a positive effect on healthcare outcomes either directly or through improvement of the quality of medical leadership (Frich et al., 2015; Geerts et al., 2020; Kumar et al., 2020; Lyons et al., 2018, 2020; Sultan et al., 2019).

The number of studies of medical leadership development programmes and of their impact on healthcare outcomes is growing, and there are indicators that the quality of research methods is improving (Frich et al., 2015; Geerts et al., 2020; Lyons et al., 2020). However, our understanding of medical leadership development is still limited (Clark & Armit, 2008; Leslie et al., 2005; Straus et al., 2013; West et al., 2015), and the availability of leadership development for both medical students and doctors remains low (Abbas et al., 2011; Baird et al., 2018; Ireri et al., 2011; Webb et al., 2014).

A range of approaches have been taken to researching medical leadership development programmes. Much initial research was conducted within quantitative paradigms, with variable focus on individuals and on context. As shown in the systematic reviews presented in Chapter Five and Chapter Ten, mixed methods approaches are now becoming more well-established in medical leadership development research. While research into medical leadership development programmes is much younger than research into leadership development broadly, medical leadership development research has gained considerable attention in recent years, and continues to develop as a field.

Evidence regarding leadership development for medical students and for doctors will be explored through systematic reviews in Chapter Four (A Systematic Review of Leadership Training in Medical Schools) and Chapter Ten (Evidence-Based Medical Leadership Development: A Systematic Review).

Chapter Conclusion

In this chapter I have briefly introduced some of the historical and contemporary approaches to leadership and presented a pragmatic definition of medical leadership for the purpose of this thesis. I have introduced the discourse around medical leadership and the lack of research into impact medical leadership development programmes may

have on healthcare outcomes. In Chapter Three I will explore methodologies which have been used in leadership research and in medical education and describe the methodologies I have chosen to use in this thesis.

Chapter Three - Methods

Chapter Introduction

In Chapter Two I reviewed existing literature around leadership, medical leadership and medical leadership development. I established the need for further research regarding the effectiveness of medical leadership development, particularly on healthcare outcomes.

In Chapter Three, I will explain the ontology and epistemology within which my research falls. I will introduce some of the methodologies which have informed existing research into leadership development, training programme evaluations and medical education. I will explain that there is not yet a consensus as to the most appropriate approaches to evaluation of leadership development in general or specifically in medical education. I will explain how I will, therefore, combine some of the insights from each of these areas, and justify the use of evidence synthesis, case studies, competency-based questionnaires, and interviews and focus groups.

As elegantly articulated by Yukl, "the purpose of the research should dictate the methodology and choice of samples, not the other way around" (Yukl, 1989).

My main research question for this thesis as proposed in Chapter One is, "***How can evaluation of medical leadership development programmes be improved?"***.

I will describe in this chapter ontological and epistemological challenges with researching medical leadership development. I will then explore existing frameworks, methodologies and methods that have been employed in evaluation of training programmes, leadership development and medical education, before explaining the methodological decisions I have made in my research.

Ontology

For the purpose of this thesis, ontology, the study of the nature of reality or existence, broadly can take objectivist or subjectivist perspectives.

An objectivist ontology holds that reality is independent of humans' perceptions, whereas a subjectivist ontology holds that reality is constructed by the observer, and therefore may differ according to different perspectives and in different contexts (Denscombe, 2010; Searle, 2006).

Because leadership is an inherently social phenomenon (as demonstrated by Stogdill's basic conditions for leadership, see chapter three), approaches that seek to take a purely objectivist or purely subjectivist stance cause ontological challenges for the study of leadership development.

Research in medicine is so firmly rooted in an objectivist ontology that it is difficult to find literature exploring ontological perspectives in medicine. Despite historically having been seen as an art, modern doctors are routinely "enculturated" into objectivist paradigms of science, and taught to search for universal truths (Wilson, 2000).

Taking such an objectivist stance in leadership development research would suppose that leadership principles are generic or universal and transferrable to all situations and contexts. However, there is also clear evidence that leadership is inherently linked to the individuals and context involved (for example, Dopson et al., 2016; West et al., 2015), which suggests that leadership development cannot be understood through a purely objectivist perspective. There nonetheless remain empirically observed associations between medical leadership and improved healthcare outcomes, as explored in Chapter Two. This would suggest that leadership development cannot be understood through a purely subjectivist stance either. This ontological tension may be part of why there remains no accepted definition of leadership.

Medical leadership therefore exists in an ontological grey zone, with no shared understanding of what the nature of medical leadership is. Rather than attempting to artificially impose an ontological framework onto this research, I have held this ontological tension throughout my research. Such an approach, with a focus on the practical rather than the abstract, has been termed “pragmatism” by some research philosophers (Biesta, 2010; Denscombe, 2010; Erickson & Noonan, 2018).

Epistemology

Given the lack of a clear ontological framework within which to situate this research, it will inevitably be problematic to assume a specific epistemological standpoint.

This is not a new problem with respect to education or training research. Indeed, Kirkpatrick’s Framework, explored below, is routinely sub-sectioned into objective and subjective outcomes and a variety of epistemological approaches are considered appropriate to learn about these outcomes (Hammick et al., 2010; Kirkpatrick, 1979; Patton, 2008) (see section below). Michael Quinn Patton holds that “it is not necessary, in my opinion, to swear vows of allegiance to any single epistemological perspective.” (Patton, 2002, p. 136).

The topic of this thesis, evaluation of medical leadership development programmes, concerns developmental programmes with practical intentions. With these practical intentions in mind, I have chosen not to focus my research on clarification, definition and articulation of an ontological framework and epistemological stance. Rather, in line with Patton’s position, I have used a variety of methods selected for their appropriateness to my pragmatic research questions.

Methodology

Given my focus on conducting research that will have practical implications for the evaluation and design of medical leadership development programmes, I have chosen a utilisation-focused evaluation approach using mixed methods (Erickson & Noonan, 2018; Patton, 2002). In a utilisation-focused evaluation approach, "a focus on *intended use by intended users* undergirds and informs every design decision in the evaluation" (original emphasis) (Patton, 2002, p. 173). Such an approach does not preclude the use of any particular methodologies, nor does it require a particular ontological or epistemological stance (Erickson & Noonan, 2018; Patton, 2002).

The appropriateness of particular methods chosen when using a utilisation-focused evaluation approach is contingent on their usefulness in producing data that can inform practical decisions by the intended users (Patton, 2008). I have therefore used mixed methods in my research, as explained in more detail later in this chapter.

The starting point for utilisation-focused research into medical leadership development is necessarily the intended use of such research by decision-makers (Hannum et al., 2007). The particular decisions and decision-makers that are the focus of the research in this thesis concern:

- Design of programmes (programme faculty);
- Funding for programmes (commissioners);
- Integration of programmes into hospitals and training programmes (organisations);
- Participating in programmes (participants).

The early stages of this research were therefore informed by informal interviews and observations with programmes, participants, faculty and commissioners.

It became clear in the course of this research that despite this early scoping work, I did not have a sufficient understanding of the data that these key decision-makers use to make decisions about programmes. Chapter Nine therefore returns to the utilisation-focused evaluation approach and focuses specifically on an evaluation of the uses and expectations that key decision-makers have for evaluation, and the data which can support their decisions.

Critiquing Utilisation-Focused Evaluation

There are four main areas where utilisation-focused evaluation has been criticised, as eloquently articulated by Michael Scriven in the Claremont Debates (Donaldson et al., 2010).

Firstly, a utilisation-focused evaluation approach starts with clarification regarding the intended use of the evaluation by the intended users. Scriven points out when an evaluation is designed based on the professed intended use by the intended users, the quality and utility of that evaluation relies on the ability of those parties to articulate their planned use, and also relies on the stability and transparency of the articulated uses. He cynically points out that “the minute they find out the true cost of everything, they tend to renege on their intended use, or the intended users, or both” (Donaldson et al., 2010, p. 22). While there is undoubtably truth in this argument, it does not negate the point of designing an evaluation with the expressed intended use by intended users in mind. Rather, as Michael Quinn Patton points out, it is essential that intended use by intended users is revisited throughout the evaluation process, and that intended users feel some ownership of the evaluation (Donaldson et al., 2010; Patton, 2008). Taking this viewpoint, it is particularly appropriate that in Chapter Nine, I return to stakeholder interviews to re-establish their intended uses and expectations for evaluation.

Secondly, Scriven argues that the “intended users” focus of utilisation-focused evaluation risks losing focus on the impact of a programme on participants by focusing rather on the intentions and expectations of those who commission an evaluation (Donaldson et al., 2010, p. 23). While an important concern, my research takes the approach suggested by Patton (Donaldson et al., 2010; Patton, 2002, 2008, 2018) that programme participants should be included as stakeholders and as intended users of evaluations. Because the participants in the case studies I introduce below volunteered to participate in these programmes, it makes sense to include them as key decision makers (“should I voluntarily participate in this programme?”). As shown in both Chapter Five (Lyons et al., 2018) and in Chapter Ten (Lyons et al., 2020), most medical leadership development programmes have evaluated programmes where participants either volunteer or apply for the programme, and are thus explicitly decision-makers with respect to those programmes. Patton’s view is that even when programmes are mandatory, participants should be considered as intended users (Donaldson et al., 2010).

Thirdly, because utilisation-focused evaluation necessitates an ongoing relationship between the users of an evaluation and the evaluator, Scriven argues that the approach risks an evaluator losing their externality by influencing the development of a programme. He suggests that if an evaluator is no longer seen as external, they may lose credibility and thus their potential to have an impact on decisions regarding a programme (Donaldson et al., 2010, p. 30). This point is particularly relevant when there are high-stakes decisions being made on the basis of summative evaluations. As Patton points out, where the purpose of evaluation is formative or developmental, there are clear benefits to an evaluator having ongoing communication with decision-makers. In particular, this ongoing communication can help to ensure that decision-makers understand the data that is collected and analysed in a utilisation-focused evaluation,

and that there is clear alignment between a decision-maker's needs and expectations for the evaluation and the information the evaluator provides them.

Finally, there are a range of factors that impact whether the recommendations from an evaluation are used. Scriven argues that it is therefore unreasonable to judge the effectiveness of an evaluation based on whether the results are used by the intended users, and that evaluations should be judged instead on their potential useability (not their actual use) (Donaldson et al., 2010, p. 31). There is inarguably some merit to this point. Politics, finances, personal relationships and a wide variety of other factors can certainly impact whether the results of an evaluation are used to make decisions. Patton argues, however, that the indirect relationship between an evaluation and the use of the evaluation highlights the importance of early and continued engagement between an evaluator and the intended users, so that the evaluator can better investigate factors that may impact the use of an evaluation (Donaldson et al., 2010; Patton, 2008). In line with Scriven's suggestion, in Chapters Six, Seven and Eight I have critiqued the evaluation methods introduced based on whether they were able to answer the questions they were designed to answer, rather than on whether the evaluation methods were used by the intended users. I have also retained a focus on the use of those evaluations. In Chapters Nine and Eleven I re-evaluate the intended use of these evaluations by the intended users, and suggest further areas for evaluation and for research.

Evaluating training programmes: frameworks

When selecting which methods and methodologies were appropriate for different aspects of this research, it has been helpful to explore established frameworks for outcomes and evaluations of training programmes. In this section, I will explore and critique two frameworks for evaluation of training programmes. I will first focus on Kirkpatrick's Framework for evaluation of training programmes (Kirkpatrick, 1967, 1979) and then on Avolio's adapted framework for evaluation of leadership development

programmes (Avolio, Reichard, et al., 2009). These frameworks helped to inform my methodology and methods decisions.

Kirkpatrick's Framework

Kirkpatrick's Framework, developed in 1960, remains the most popular approach to evaluation of training programmes in organisations (Arthur et al., 2003; Bates, 2004; Powell & Yalcin, 2010; Saks & Burke, 2012). Kirkpatrick framed the outcomes of training programmes into a four-part hierarchy, as shown below (Table 2)

Level	**Descriptions (Kirkpatrick, 1967)**
Level One: Reaction	How well did [learners] like the programme?
Level Two: Learning	What principles, facts and techniques were learned?
Level Three: Behaviour	What changes in job behaviour resulted from the program?
Level Four: Results	What were the tangible results of the program in terms of reduced cost, improved quality, improved quantity, etc?

Table 2: Kirkpatrick's Training Programme Outcome Framework

Kirkpatrick emphasized that these steps were a guide for evaluation rather than a fixed protocol (Kirkpatrick, 1967, 1979). A range of adaptations of Kirkpatrick's steps have been developed for evaluation use across organisations, including in healthcare and in leadership development (see Avolio's Categories below) (Barr, 1999; Hammick et al., 2010; Roland, 2015; Yardley & Dornan, 2012).

Five main criticisms have been repeatedly levelled against the use of Kirkpatrick's Framework in healthcare and elsewhere. These are that the model is incomplete, neglects the effect of context, assumes causality, promotes a fixed hierarchical view of the relative importance of information, and that evaluation at levels three and four are challenging (Bates, 2004; Moreau, 2017; Yardley & Dornan, 2012). I will introduce each of these criticisms below and explore their relevance to my research.

Critiques of Kirkpatrick's Framework

Incompleteness

The first and most important criticism of Kirkpatrick's Framework is that it does not fully encompass the possible outcomes of a programme. Training programmes may have a range of tangible and intangible outcomes. As well as observable changes in behaviour, for example, a training programme could result in transformative but less observable changes to the way a person thinks. Some researchers caution that the wide acceptance of Kirkpatrick's Framework and the comparative ease of evaluating tangible and quantifiable outcomes could lead to neglect of important intangible and soft outcomes (for example, Hannum et al., 2007; Yardley & Dornan, 2012)

Focus on outcomes encapsulated by Kirkpatrick's Framework may well have influenced the focus of early leadership development research on quantitative outcomes. As mixed methodology approaches to evaluation of development programmes become more prevalent, the incompleteness of Kirkpatrick's Framework is likely to become both more noticeable and less important. Qualitative approaches, which inductively evaluate a horizon of outcomes from a programme, are not limited by the frameworks held implicitly and explicitly by researchers and evaluators (Hannum et al., 2007). They are therefore able to provide insights that are not predicted by researchers, nor limited to the four levels of Kirkpatrick's Framework.

Neglect of context

Kirkpatrick's Framework focuses on the outcomes of a programme rather than the design of the programme or the experience of the individual, and neglects context. It is increasingly recognised that learning is complex and takes place in open systems where innumerable interactions between learners and their environments affect the outcomes of a programme (Hannum et al., 2007; Patton, 2002). It is therefore generally not possible

to draw causal inferences from differences in evaluations without further information (Bates, 2004; Roland, 2015; Yardley & Dornan, 2012).

If we are to synthesise learning across programmes it is vital that evaluators and researchers explore and communicate context for programmes, and that educators conduct needs assessments in their programme development (Arthur et al., 2003)

Assumption of causality

Bates (2004) highlights that a range of studies have failed to confirm causal linkages between Kirkpatrick's levels. It should be remembered that absence of evidence is not evidence of absence.

From an educational point of view there are at least theoretical reasons to expect that participant changes in knowledge, skills or attitudes are necessary for a change in behaviour to occur, and a change in behaviour is necessary for tangible results to occur (Arthur et al., 2003). It is less clear whether a positive reaction to a programme is a necessary condition for learning to take place, and research has generally failed to show reliable relationships between reaction and the other three levels (Arthur et al., 2003).

The presence or absence of causality could be seen as a moot point, however; it is an accepted principle that an evaluation should seek to evaluate outcomes at as high a level as feasible and at as many levels as is feasible (Hammick et al., 2010; Hannum et al., 2007; Saks & Burke, 2012).

While the level of the outcome may have an impact on the usefulness of data collected for decision-makers, assuming causality between Kirkpatrick levels of outcomes would seem to only be a problem if an evaluator was to assume that positive reactions necessarily translate into tangible outcomes (Saks & Burke, 2012). In this thesis I make an explicit point of not assuming that positive reactions necessarily translate into tangible outcomes.

Relative importance of levels of outcomes

Yardley and Dornan highlight the danger in assuming that higher Kirkpatrick levels are more important than lower levels (Yardley & Dornan, 2012). Indeed, it is conceivable that holding a set hierarchy of outcomes in an evaluator's mind could lead to neglect of the first step in effective utilisation-focused evaluation research (collaboration with stakeholders to establish clear goals and outcomes for a programme) (Hannum et al., 2007; Patton, 2008).

Programmes can have a variety of intended outcomes, as will be explored in Chapter Ten. While traditionally there has been emphasis placed on tangible organisational outcomes (such as financial impact, improved productivity, improved patient outcomes), it is important to establish the intended outcomes of a programme, and the outcomes from which decisions will be made.

Furthermore, a focus on evaluating tangible results could risk short-term focus, rather than focus on transformative developmental changes that might result in indirect but more impactful outcomes (Grove et al., 2005; Hannum et al., 2007).

Evaluation at Kirkpatrick Level Three and Four is challenging

Across a range of industries including healthcare there is a preponderance of research and evaluation that focuses on Kirkpatrick Level One reactions (Kennedy et al., 2014; Saks & Burke, 2012).

Yardley and Dornan caution that if reviews of research reject evaluations that concern only Level One outcomes, important learning from those evaluations might be neglected (Yardley & Dornan, 2012). It could also be pointed out that an evaluation that is limited solely to exploring the reaction of participants to a programme, without exploring learning or application of that learning into their work, is certainly missing significant opportunity

for insights. As quoted by Arthur et al., “reaction measures are not a suitable surrogate for other indexes of training effectiveness” (Arthur et al., 2003).

Summary of critiques of Kirkpatrick’s Framework

It is important to bear in mind the assumptions and limitations of any framework when applying it into research. Kirkpatrick’s remains the main framework to guide evaluation, despite critiques over the years (Cook, 2010; Frye & Hemmer, 2012; Haji et al., 2013; Yardley & Dornan, 2012). Kirkpatrick’s Framework offers a helpful starting point for establishing the scope of outcomes which might be of importance in evaluation, especially when combined with modern mixed methods approaches.

Avolio’s categories

In 2009, Avolio and colleagues conducted a thorough systematic review and meta-analysis of experimental and quasi-experimental leadership impact research which “attempted to change and/or develop leadership” (Avolio, Reichard, et al., 2009). Quasi-experimental research was defined as studies which did not use random assignment.

Avolio’s team categorised outcomes into affective, behavioural, cognitive and organisational performance:

- **Affective** outcomes regarded a participant’s liking, satisfaction or enjoyment of the intervention, and fit with Kirkpatrick’s first level, “reaction”.
- **Behavioural** outcomes included leader emergence, participation, or performance of certain behaviours, and fit well with Kirkpatrick’s third level, “behaviour”.
- **Organizational** performance included profit or productivity increases, and fit with Kirkpatrick’s fourth level, “results”.
- **Cognitive** outcomes were described by Avolio’s group as concerning “idea generation or confidence”.

Many cognitive outcomes within this description would fit within Kirkpatrick's second level, "learning". A range of outcomes which are more difficult to fit into Kirkpatrick's Framework, such as changes in thinking, would fit within Avolio's cognitive outcome category.

By including cognitive outcomes, and by not explicitly arranging the affective, behavioural, cognitive and organisational outcomes in a hierarchical framework, Avolio et al. (2009) effectively sidestepped some of the criticisms of Kirkpatrick's Framework as explored earlier in this chapter. His team nonetheless lamented the same difficulty categorising outcomes that we found in our systematic reviews of leadership development in medical schools (Chapter Five) (Lyons et al., 2018) and for doctors (Chapter Ten) (Lyons et al., 2020). This difficulty has also been commented on in other reviews which have attempted to categorise outcomes into a pre-existing framework (Frich et al., 2015; Geerts et al., 2020; Onyura et al., 2019; Rosenman et al., 2014; Sadowski et al., 2018; Steinert et al., 2012; Sultan et al., 2019; Webb et al., 2014).

Evaluating leadership development programmes: methodologies and methods

As highlighted in Chapters One and Two, leadership development is not new, and nor is evaluation of leadership development initiatives. It is nonetheless still widely accepted within the literature that there remains a lack of research evaluating outcomes from leadership development programmes, and a lack of suitable evaluation instruments (Avolio, Reichard, et al., 2009; Black & Earnest, 2009; Hannum & Craig, 2010) (see Chapter Two).

The Centre for Creative Leadership (CCL), set up in 1970, has published a handbook focused entirely on evaluation of leadership development (Hannum et al., 2007), with input from more than 30 experienced evaluators. While this handbook offers a range of

valuable insights for a utilisation-focused evaluation approach, it falls short of establishing optimal methodologies for evaluation of leadership development.

Indeed, the CCL handbook brings to the forefront the often implicit and loosely defined intentions of programme sponsors and faculty. “All too often, leadership development initiatives are implemented without a clearly stated set of objectives or outcomes” (Hannum et al., 2007, p. 24).

Mixed-Methodologies

Established best practice in evaluation of leadership development programmes starts with clarification of the intended goals for a programme (Edmonstone, 2013; Hannum et al., 2007; Patton, 2008). While the dominant research approaches in the development of the theories of leadership explored in Chapter Two were experimental and quantitative in nature, it is now becoming generally accepted that mixed methodology approaches offer richness and insight that is not captured within purely quantitative approaches (Black & Earnest, 2009; Hannum, 2004; Hannum et al., 2007; Leskiw & Singh, 2007; O’Neill, 2016; Orvis & Ratwani, 2010; Parry & Bryman, 2013; Patton, 2008; Russon & Reinelt, 2004; Yukl, 1989).

There remain a range of opinions as to which methodologies should be included within leadership development research. Experts point out that leadership development is complex, operating in open systems where participants may be influenced by a range of interactions both within and without programmes (Day et al., 2014; Hannum et al., 2007; Patton, 2008). Due to the fact that leadership development is a dynamic process involving concurrent and varied interactions over time (Day et al., 2014), it has been argued that “isolation of any one particular aspect to allow the attribution of cause and effect is by definition problematic” (Edmonstone, 2013).

Bespoke methods

Indirect causal relationships combined with variation in stakeholder expectations and variation in definitions of leadership itself leads many experts to suggest simply that evaluation methods be developed collaboratively with stakeholders, sponsors and participants, and that attention be paid to context (Day et al., 2014; Hannum et al., 2007; Patton, 2008).

Bespoke evaluation methods offer clear benefits in terms of alignment of evaluation efforts with the needs of programme decision-makers, particularly if a broad framework is used to guide those evaluations.

It does, however, become difficult to create knowledge across programmes if participants and stakeholders for every programme develop their own outcome measures and assess their programme formatively and summatively against those bespoke measures only. Indeed, after emphasizing the need for bespoke evaluation methods which take into account stakeholder requirements, the CCL handbook concludes that one of the key areas for further research is “what program components are most strongly related to which program outcomes” (Hannum et al., 2007, p. 562).

Existing quantitative methodologies

While Avolio’s review (2009) was extensive, it did not report evaluation methodologies used by the included studies in detail. A later study led by Jessica Dinh (2014) produced a summary of methodologies used across 542 leadership articles published in top-tier management journals (see Table 3).

Dinh likewise stops short of analysing the actual methods used in these articles, though she notes some of the most commonly used methods across different methodological approaches included questionnaires and surveys (Dinh et al., 2014).

Methodology	Number (%) of studies

Field survey (primary data from participants)*	365 (67%)
Field survey (secondary data from archives)	122 (23%)
Laboratory experiment (tasks outside context)	98 (18%)
Content analysis (quantitative analysis of interviews)	74 (14%)
Field experiment (applied research in context)*	23 (4%)
Observation*	6 (1%)
Diary study/experiential sampling	5 (<1%)
Experimental simulation	5 (<1%)
Computer simulation	4 (<1%)
Judgment task	3 (<1%)
Methodology study (for example, psychometric methods)	3 (<1%)

Table 3: Prominent quantitative methodologies for evaluating leadership. Adapted from Dinh et al (2014). Review methodologies have been removed. *indicates methodologies used in this thesis.

Gardener and colleagues provide instructive exemplars for each of these categories, and recommend reduction in reliance on purely retrospective survey measures; increase in the use of multiple methods; increase in the use of controlled experiments in the laboratory and in the field; increase in the number of follow-up studies which research the same programme or context (Gardner et al., 2010).

In addition, there is a growing push (Day et al., 2014) towards incorporation of new methods into the leadership development evaluation toolbox, including social network analysis (Hoppe & Reinelt, 2010) and Q-methodology (Militello & Benham, 2010), and towards increased attention to Return On Investment (ROI) (Avolio et al., 2010).

Russon and colleagues point out that questionnaires and surveys are common because they are cheap to employ, but they limit the potential for triangulation and do not offer the richness of data of interviews or focus groups (Russon & Reinelt, 2004). Hannum offers a selection of methods which can be applied within many of Dinh's methodologies, and Leskiw suggests the main outcome level (individual, team or organisation) of each of these methods (Hannum, 2004; Leskiw & Singh, 2007) (see Table 4).

Method	Outcome level	Kirkpatrick Level	Avolio Category	Description
Daily Evaluations*	Individual	Reaction	Affect	Daily evaluations of participant experiences (that is,, on a programme)
End-of-initiative evaluations*	Individual	Reaction	Affect Cognition	Post-programme or post-component evaluations concerning reaction to programme and intention to apply learning
Interviews*	Individual	Reaction Learning Behaviour Results	Affect Behaviour Cognition Organisation	Can be used pre, during or post programme for a variety of purposes which include identification of training needs, participant experiences on a programme, participant application of learning, and contextualisation of survey results.
Learning surveys*	Individual	Learning	Cognition	These can take the form of direct examinations of knowledge before and after a programme, or indirect self-assessments.
Change surveys*	Individual	Behaviour	Behaviour	Most effective when focused on behavioural changes, a change survey is usually administered both before and after a programme, and asks participants to self-assess their behaviour at different time points.
Behavioural observation*	Individual	Behaviour	Behaviour	Behavioural observations can produce both qualitative and quantitative data, by observing participants (in or out of their usual context).
Focus groups*	Team	Reaction Learning Behaviour Results	Affect Behaviour Cognition Organisation	Focus groups obtain qualitative information from a group of individuals involved in a programme. As with interviews, these can be conducted at various time-points.
Return-on-investment (ROI)	Organisation	Results	Organisation	ROI usually refers to the financial cost-benefit analysis of a programme, though it can concern non-financial returns. ROI isolates expected organisational results and calculates the financial impact these represent. ROI is challenging to calculate for programmes with predominantly intangible expected outcomes.
Workplace Statistics	Organisation	Behaviour Results	Behaviour Organisation	Many organisations routinely collect data regarding absences, conflicts, complaints, and such. Where a programme is specifically expected to affect these outcomes, collection of workplace statistics can be an excellent evaluation method. In a hospital, these could include patient outcome statistics.
Document Analysis*	Organisation	Learning Behaviour	Behaviour Cognition	Document analysis is a relatively wide set of methods involving the analysis of written information, and can include written vignettes and

		Results	Organisation	reflections, financial accounts, patient records, performance, communications, curriculum vitae analysis and written forms of many of the other methods.
Appraisal of customer satisfaction	Organisation	Results	Organisation	Customer satisfaction is a common endpoint outcome that organisations routinely collect. These outcomes are often collected by surveys but could also use interviews or focus groups.
Climate survey retests	Organisation	Results	Organisation	Many organisations conduct annual climate assessments. Where an intervention can reasonably be expected to affect climate, the use of before and after programme climate assessments may offer an indication of organisational impact, though there are clearly a range of mediating factors.
Culture survey retests	Organisation	Results	Organisation	Culture, or the behaviours and values that pervade an organisation, and changes slowly. Where an initiative is specifically designed to affect culture, culture can be assessed by a range of methods before and after a programme, usually over a time-frame of years.

Table 4: Methods for evaluation, compiled and adapted from Hannum (2004).
Outcome levels from Leskiw (2007). Indicative Kirkpatrick and Avolio categories added. *indicates methods used in this thesis.

Weaknesses of existing approaches to evaluation of leadership development

Lack of longitudinal research

There is broad agreement that there is too much use of cross-sectional evaluation designs in leadership development and there is a need for more longitudinal studies with measurements at two or more timepoints (Avolio, Reichard, et al., 2009; Day et al., 2014; Dinh et al., 2014; Russon & Reinelt, 2004).

Focus on indirect outcomes (job performance)

Day and colleagues caution that solely focusing on job performance risks too much contamination by factors other than leadership, and suggest that research should focus on identifying and tracking markers or proxies of development that extend beyond job performance (Day et al., 2014).

Reliance on participant self-assessments

Russon highlights an over-reliance on data collected from programme participants (Russon & Reinelt, 2004), and Solansky quotes a number of studies which caution against reliance on self-reports "which are prone to many kinds of response bias" (Solansky, 2010). She suggests that the use of multi-source feedback can help to mitigate the reliance on self-reports, as they "represent a more rigorous methodology for the evaluation of leadership development programs than self-report only data by gathering data from multiple sources" (Solansky, 2010).

Conflation of leader development with leadership development and neglect of context

Dalakoura (2010) highlights conflation of leader development and leadership development as a problem, particularly given the way in which leadership effectiveness

is mediated by context. She cautions that leadership development relies on the "expansion of the collective capacity of the people of the organisation to effectively engage in leadership roles and processes" (Dalakoura, 2010). This emphasis is echoed in a number of other reviews, including in healthcare (Bohmer, 2012; Day, 2000).

Evaluating medical education: framework and methodologies

The medical profession has long evaluated and examined trainees as part of certifying them ready to practice medicine or to practice as a specialty. The concept of evaluating medical education itself, nonetheless, is relatively new. Since Harden and colleagues produced the first 'Best Evidence in Medical Education' (BEME) guide in 1999 (Harden, Grant, & Buck, 1999) there have been a variety of advancements, however, medical education is certainly no further ahead than leadership development in terms of evaluation. For the most part, the methodologies that have been used to evaluate medical education have come from the same traditions as the methodologies for evaluating leadership development.

Harden suggested the use of Kirkpatrick's Framework to guide evaluation of medical education in 1999 (Harden, Grant, & Buck, 1999).

As focus has shifted in medicine from proving education initiatives to improving initiatives, from summative to formative and explorative, a greater emphasis has arisen for mixed methods, in parallel with the shift seen in leadership development (Cook et al., 2008; Durning et al., 2007; Frye & Hemmer, 2012; Haji et al., 2013; Yardley & Dornan, 2012). There is also a greater emphasis on producing useful data for programme decisions, and for incorporating stakeholder consultation and collaboration throughout the evaluation process (Cook, 2010; Lauer, 2018; Vassar et al., 2010).

Frye and Hemmer's 2012 review still suggested interviews, surveys, records analysis (test results, learner performance), document reviews, and focus groups as the main methods to be selected from during an evaluation (Frye & Hemmer, 2012).

Continual assessment of student and trainee progress is a persistent and important part of medical education (Norcini & McKinley, 2007). There are multiple points for alignment of educational evaluations with existing assessment protocols, which (if well-chosen), can be excellent indicators of learning and change in behaviour. Durning, Hemmer and Pangaro have compiled an excellent indicative list of some of these assessment opportunities in the setting of US undergraduate and graduate medical education (Durning et al., 2007).

Cook and colleagues, explicating standards for medical education papers for the Journal of General Internal Medicine, note that "much medical education research relies on a survey as a means of collecting data" (Cook et al., 2008). As will be noted later, this statement is true also in leadership development research in medical education (Frich et al., 2015; Lyons et al., 2018, 2020; Webb et al., 2014) (see Chapter Five and Chapter Ten). Surveys can be a useful tool if carefully designed and validated, but their utility is limited if their results cannot be confidently linked with the outcomes they are intended to measure.

Methodological conclusions

In summary, there are no agreed methods for evaluating medical leadership development programmes. Rather, there are some key principles which arise across training evaluation frameworks, leadership development evaluation and medical education evaluation that have informed my methodological decisions.

First, there are a wide range of methods and methodologies that can be employed to collect useful information. The use of self-reported single-method surveys is shown to be

insufficient and the emphasis in evaluation is increasingly on the use of multiple methods. The use of multiple methods helps evaluators to collect richer data regarding a programme of development, which can better inform decision-makers, including regulators, funders, faculty and participants.

Second, while it is helpful to ground an evaluation in an accepted framework such as Kirkpatrick's, it is also important to maintain openness to formative outcomes, to outcomes which do not fit into the framework and to the context an evaluation is conducted in, and it is important to not assume linkages between levels of outcome or indeed between measurements and outcomes.

Methodology for this thesis

In alignment with the conclusions above, I have pragmatically adopted a utilisation-focused evaluation approach for this research. I have situated both my reviews and my primary research within Kirkpatrick's Framework, while remaining cognizant of the importance of context and rich data.

As outlined in Table 3 above, my methodologies for this thesis include systematic reviews, field surveys, field-experiments and observations. Within those methodologies I have used mixed methods including documentary analysis; case studies; competency-based self-assessment questionnaires; surveys; interviews and focus groups.

As I briefly explained in Chapter One, throughout the work I have presented in this thesis my intention was to produce research with potential for impact. I therefore adjusted the setting of my research in order to increase my potential for impact. Initially my research focused on medical leadership development for medical schools in New Zealand and Australia (Chapter Four) and medical schools internationally (Chapter Five). As I moved into development of evaluation methods, I changed the setting of my

research to focus on medical leadership development in the NHS (Chapters Six, Seven, Eight and Nine).

Of the four countries that have had the strongest focus on medical leadership over the last two decades (UK, United States, Canada and Denmark) (Spurgeon et al., 2011), I had best access to medical leadership development case studies in the UK (see Chapter Three). This was therefore the most appropriate setting for me to conduct research with potential for impact.

In order to tie my research together, I finally returned to an internationally setting for the systematic review presented in Chapter Ten (Lyons et al., 2020).

Case studies

In this thesis I have included case studies of five programme cohorts over three programmes based in Oxford University Hospitals. The case studies used in this research are named below, and described in more detail in Appendix Six – Case study descriptions:

1. *FLP – Oxford Future Leaders Programme (2016-2017):* A 12-month longitudinal leadership development programme for newly qualified medical consultants, delivered on-site;
2. *ELP1 – Oxford Emerging Leaders Programme (Cohort 1):* A 5-month longitudinal leadership development programme for junior doctors on representative committees, delivered on-site;
3. *ELP2 – Oxford Emerging Leaders Programme (Cohort 2):* An 8-month longitudinal multidisciplinary leadership development programme for medical trainees and other healthcare professionals at Oxford University hospitals, delivered on-site;
4. *ELP3 – Oxford Emerging Leaders Programme (Cohort 3):* A second 8-month longitudinal multidisciplinary leadership development programme for medical

trainees and other healthcare professionals at Oxford University hospitals, delivered off-site;

5. *IMPACT - Improving Performance and Care in Teams:* A 12-month longitudinal leadership development programme for senior management triumvirates at Oxford University Hospitals (Matrons, Clinical Directors, Operational Service Managers), delivered off-site.

These three programmes (the FLP, ELP and IMPACT) made up the entirety of the longitudinal leadership development programmes which included doctors and were based in Oxford University Hospitals.

Individual doctors at Oxford University Hospitals also participated in other leadership development programmes, including online programmes run by the NHS Leadership Academy. These external programmes include a wide range of participants from different hospitals. They would have been difficult to gain appropriate access to study, and the wide range of contexts that the participants experienced would have introduced a considerable number of additional confounding factors, so they were not included as case studies in this thesis.

The participants in the included case studies spanned a range of professions and levels of experience. Each was designed to enhance experiential learning, and the setting for all programmes was the same hospital. By focusing on case studies that shared an organisational setting, it was possible to gain a greater understanding of how the context impacted the effectiveness and the evaluation of these programmes.

Navigating my personal involvement in the case studies

I was involved as an external evaluator and observer in both the FLP and IMPACT programmes. These constituted the clearest delineation of my involvement, though it should be noted that one of the FLP faculty also supervised my DPhil.

I was involved in setting up the Emerging Leaders Programmes (ELP), including design and delivery of the programmes. In the course of my research surrounding the FLP I was asked by the Director of Medical Education at Oxford University Hospitals, Professor Peter Sullivan, if I would be interested in setting up a leadership development programme for medical trainees, based on the FLP. Given the difficulty of obtaining access to relevant cases to study, this presented me with an opportunity and a dilemma.

Because I was involved in the design of the ELP, I was able to build evaluation into the programme design, and I was able to trial different evaluation methods in later programmes. It also meant that I was intimately familiar with the material of the programme, and with the context in which it was situated. As both Hannum and Patton point out, understanding of context is a considerable boon for internally-placed researchers (Hannum et al., 2007; Patton, 2008).

My involvement in the design and delivery also presented a dilemma for me in terms of evaluation.

While I was not paid for my role in this programme, I voluntarily invested considerable time into design and delivery. As participants were aware of my personal investment in providing them with a learning opportunity, it is possible they may have been affected to a greater extent by social desirability biases when completing questionnaires (see discussion section of Chapter Seven for further exploration of biases). To help ameliorate these biases, post-programme interviews with ELP programme faculty and participants were conducted by a research assistant, with the aim of participants feeling more comfortable offering critical and negative comments about the programme.

I am also cognizant of the possibility of me personally being affected more strongly by effort justification bias than were I to be an entirely external evaluator. With a view to trying to mitigate the effects of my close involvement with the ELPs, I had a strong focus

in my analysis of both quantitative and qualitative data on seeking disconfirming evidence and opportunities for improvement.

Chapter Conclusion

In this chapter I have explained my philosophical and practical approach. I have reviewed existing methodological approaches from evaluations of training, of leadership development, and of medical education. These have formed the areas in which my research sits (Figure 2).

I have outlined my approach as utilisation-focused evaluation and have explained my use of multiple methodologies and methods across the chapters of this thesis in order to collect the most relevant and useful data for each component. Finally, I have explained the transition in focus from medical leadership development for medical students in New Zealand and Australia through to medical leadership development for doctors in the United Kingdom, and I have briefly described the case studies which will form the basis for Chapters Six, Seven, Eight and Ten.

The specific methods I have used are described in more detail in the relevant chapter.

In Part II I will focus on leadership development for medical students, and I will present a needs analysis of leadership development in medical schools (Chapter Four) and a systematic review of medical leadership training (Chapter Five), which I conducted at the beginning of my doctoral research to establish the current state of the literature and to situate my research going forward.

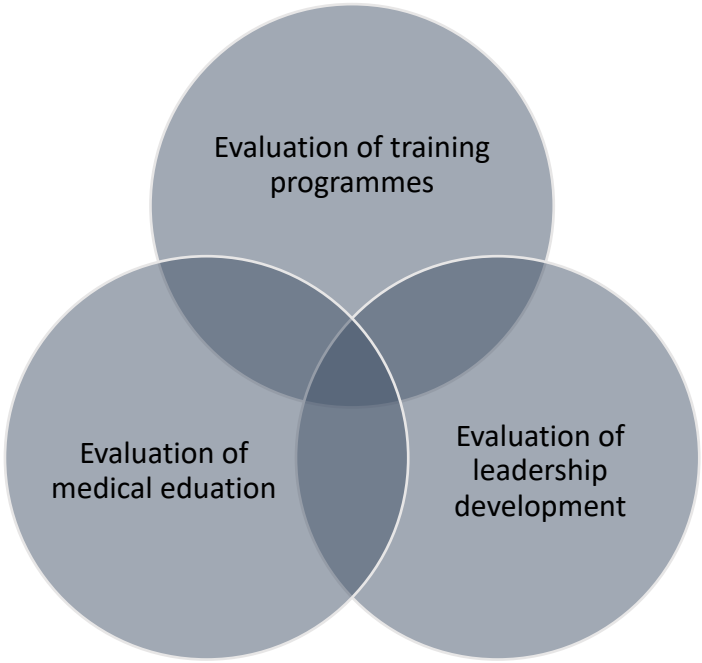


Figure 2: Situation of my research within the three methodological literatures explored in Chapter Three.
My research, evaluation of medical leadership development programmes, sits at the convergence of the three circles.

PART II:

Leadership Development in

Medical Schools

This section is focused on leadership development for medical students

Chapter Four – Needs Analysis of Leadership Development in Australian and New Zealand Medical Schools

Note on publication of Chapter Four

The research I have presented in this chapter was first published in BMJ Leader as Lyons, O., McHardy, K., Bagg, W., & Wilkinson, T. (2019). Leadership development in New Zealand and Australian medical schools: Needs analysis. *BMJ Leader*, *3*(4), 123–128. https://doi.org/10.1136/leader-2019-000147. In Chapter Four I have adapted this published paper to make it compliant with departmental regulations for including published work in a doctoral thesis, and to make it clear how this work relates to other chapters in this thesis.

Contribution statement

As the lead author for this paper, I conducted all aspects of the research and writing, including study design, mapping, matching, drafting, editing and approval of the final manuscript. For the purpose of publication, other members of this project team (Karina McHardy, Warwick Bagg, and Tim Wilkinson) independently completed parts of the research in parallel, as outlined in the text below. All co-authors approved the final manuscript, and all have approved the inclusion of this research in my thesis.

Chapter Introduction

As introduced in Chapters One, Two and Three, my research began with a focus on medical leadership development for medical students in New Zealand and Australia,

before I made a pragmatic shift to a focus on medical leadership development for doctors in the United Kingdom.

Medical councils worldwide have begun to include effective leadership as a core competency expected of health professionals (Australian Medical Council, 2012; General Medical Council, 2018; Medical Council of New Zealand, 2014). In order to establish whether current curriculum standards effectively meet leadership development standards already, I conducted a needs analysis of the Australian and New Zealand medical curriculum standards. In this chapter, I describe the process and the outcomes of this needs analysis, conducted in collaboration with a small group of colleagues based in New Zealand.

Background to the needs analysis

Despite becoming an explicit part of graduating doctors' roles it remains unclear when leadership competencies are best learnt (Jefferies et al., 2016). Some aspects of leadership undoubtedly require longer experience and greater seniority to enable practical application and experimentation, as indicated in the Medical Leadership Competency Framework (MLCF) (NHS Institute for Innovation and Improvement, 2010). Many components of medical leadership, however, could be introduced in pre-qualification medical curricula, as outlined in the MLCF Guidance for Undergraduate Medical Education (Spurgeon & Down, 2010).

Ideally, learning in any curriculum area should be 'just in time', whereby it is seen to be relevant to practice, relates to recent or imminent experience and has opportunities to apply what is learnt. Getting this timing right is a challenge in any programme, however. There are a number of reasons this may be particularly difficult when considering leadership development. Firstly, as highlighted in Chapter Two "leadership" as a concept is not always well defined. Secondly, there are different levels of leadership

ranging from leading an institution, through to being prepared to play one's part in improving a system, leading a team, collaborating with colleagues, or simply being a good follower. The relevance and opportunities to learn leadership during primary medical training vary depending on which aspect of leadership is being referred to. As such, simply recommending that all training doctors should learn about leadership lacks the specificity to be constructive. In addition, stating that it can only be learnt after graduation is also unhelpful.

In 2015 the University of Auckland trialled a collaboration with a leadership institute to introduce a one-week long leadership development programme for senior medical students. Students struggled to engage with the leadership learning, as it seemed out of context of their "becoming a doctor". Internal evaluation suggested that students had not yet established their identity as individual medical practitioners and therefore couldn't see the relevance of learning leadership at this stage, particularly when delivered from a business context.

The Australian Medical Council (AMC) requires medical schools in Australia and New Zealand to meet a range of standards ('Graduate Outcomes'), thereby ensuring that graduating students have the basic competencies to work as Resident Medical Officers (RMOs) (Australian Medical Council, 2012). With leadership accepted as one of these core competencies, these standards should support medical schools to develop the foundations of medical leadership.

My collaborators on this project were involved with both of the medical programmes in New Zealand, neither of which have redeveloped curricula yet to comprehensively meet the leadership development needs of medical students. In addition, there seems to be limited evidence elsewhere of successful implementation. Consequently, we saw an unmet need to determine those aspects of medical leadership which are not sufficiently addressed in current curricula. In line with recent work in the United Kingdom (Peake et

al., 2018), this research was undertaken to clarify which elements of medical leadership development are present in AMC outcomes and which elements are missing, and to assess which elements might reasonably be addressed in pre-qualification curricula in New Zealand and Australia. Given the similarities between medical education in the United Kingdom, findings from this work may be useful more widely such as in the United Kingdom and other similar jurisdictions.

Aims

This project sought to establish how Australasian medical schools could deliver medical leadership development that is meaningful and appropriate in curricula which are already content laden. The three specific questions we have addressed are:

1. What are the essential components of medical leadership?
2. Are there gaps related to leadership development in current AMC Outcomes for Graduates?
3. If gaps are identified, which gaps should be addressed in medical school, and which should be addressed after qualification?

Medical programmes are diverse. Therefore, our aim was not to suggest specific ways to change curricula or how to teach leadership, but to establish which leadership components should be emphasised, and where leadership development efforts could best be focused within Australian and New Zealand medical school curricula.

Defining learning outcomes for leadership

We selected the MLCF (NHS Institute for Innovation and Improvement, 2010) as currently the most evidence-based, well evaluated, internationally accepted and freely available framework for medical leadership development in medical schools. Our initial goal had been to systematically review the published leadership frameworks and to

synthesise these into a single framework. When we conducted a literature search for evidence-based published and evaluated frameworks for leadership in healthcare, we found that few published frameworks meet these two conditions. As a broad review and synthesis had already been completed as part of the development of the MLCF, we consulted with the authors of this framework and elected to use it as the basis for our curriculum comparison.

While no formal evaluation of the effectiveness of the MLCF had been published at the time, the MLCF has already been used as a tool for measuring the effectiveness of leadership training in medicine (Frich et al., 2015). The MLCF is also well-established, recognised by the United Kingdom's General Medical Council, and recently adopted into their guidelines for medical training with a clear set of recommendations for implementation at the undergraduate (pre-qualification) level (Spurgeon & Down, 2010). In contrast, the main contender at the time of this project was LEADS (Lead Self, Engage Others, Achieve Results, Develop Coalitions and Systems Transformation) which explicitly states "LEADS capabilities are not competencies and therefore do not clearly state the knowledge, skills and attitudes required of leaders" (Fenwick & Hagge, 2007).

The Royal Australasian College of Medical Administrators (RACMA) has also produced a curriculum for medical administrators. The explicit aim of the RACMA curriculum is, however, to define the "capabilities and competencies for training, attaining and enhancing of the skills of the specialist medical administrator", not to be a framework for leadership development in medical schools (Royal Australasian College of Medical Administrators, 2011).

Methods:

In line with Thomas et al.'s Curriculum Development for Medical Education (Thomas et al., 2016), we conducted a general needs analysis of the AMC curriculum, identifying

the differences between the AMC outcomes for graduates and the MLCF competencies. This process, also known as problem identification, helps ensure that subsequent efforts to improve curricula target areas of importance.

Phase 1: Contextualising the Medical Leadership Competency Framework

The MLCF is designed for doctors through to specialisation and continued practice and does not explicate which domains can or should be fully achieved by graduation. In order to ensure MLCF domains were appropriate for the pre-qualification environment, each domain was reviewed by all four authors to determine whether it would be sufficiently relevant to undergraduate experiences to enable full achievement before qualification. We used statements from the Guidance for Undergraduate Medical Education to help guide this analysis (Spurgeon & Down, 2010). In line with the principles of the MLCF, where full achievement was not possible, we considered whether domain elements could still be introduced at a scaled level. Consensus was reached by discussion.

Phase 2: Mapping Outcomes for Graduates onto the MLCF

For each MLCF domain, I searched the AMC outcomes and matched all AMC outcomes which were similar or could contribute to the relevant MLCF domain in parallel with two other team members (Warwick Bagg and Tim Wilkinson). These three lists were compiled, allowing AMC outcomes to be matched to multiple MLCF domains where relevant. One of the MLCF lead authors (Peter Spurgeon) was consulted to add further information about the MLCF than was available in published form, and to clarify questions and points of uncertainty.

Phase 3: Assessment of match

Each author independently evaluated the degree to which AMC outcomes would satisfy MLCF domain elements. Satisfaction meant that by achieving relevant AMC outcomes, a student would also meet the requirements for MLCF domain elements. Each MLCF domain element was evaluated to determine whether it is satisfied by AMC outcomes. Where domain elements were judged in phase 1 to be more appropriate to an undergraduate level than full achievement, we evaluated whether the foundations of these “postgraduate” domains would be met by current AMC outcomes. These evaluations were collated and compared, and consensus was attained by discussion. Where there was ambiguity in the MLCF domains or their associated descriptors, the authors prioritised interpretation of the guiding principle of the MLCF domain.

Results

5/8 (63%) of undergraduate MLCF domains and 7/12 (58%) of postgraduate domains were found to be already satisfied by the AMC Outcomes for Graduates. MLCF domains relating to personal values; communication; collaboration; and teamwork were well met. However, as demonstrated in Figure 3, using the current AMC Outcomes for Graduates, most undergraduate medical students would not have sufficient opportunity to contextualise and develop elements of many leadership domains. Illustrative examples of how the matching occurred are shown in Table 5.

There was no clear 1:1 relationship between many AMC outcomes and MLCF domains. Many descriptors were either too general or too specific. Therefore, significant interpretation of AMC outcome and MLCF domain intent was required before consensus could be achieved.

When matching MLCF and AMC standards we found they were often too broad or too narrow to be useful. Several AMC domains could broadly cover leadership within their bounds, for example:

4.1 Provide care to all patients according to "Good Medical Practice: A Code of Conduct for Doctors in Australia" and "Good Medical Practice: A Guide for Doctors" in New Zealand.

4.8 Describe and respect the roles and expertise of other health care professionals and demonstrate ability to learn and work effectively as a member of an inter-professional team or other professional group.

The intention of these outcomes could cover professionalism, leadership and other aspects needed to be an effective team member. They are, however, so broad they are not practical. Conversely, several AMC outcomes lie on the edge of being highly relevant to medical leadership but are worded too narrowly.

2.11 Prescribe medications safely, ***effectively and economically*** *using objective evidence. Safely administer other therapeutic agents including fluid, electrolytes, blood products and selected inhalational agents. (emphasis added)*

1.4 Access, critically appraise, interpret and apply evidence from the medical and scientific ***literature****. (emphasis added, showing how this outcome misses audit by specifying literature, rather than systems).*

Additionally, some key elements of leadership as described in the MLCF are not yet encapsulated by AMC Outcomes for Graduates (see Table 6). Two particularly notable absences are audit and quality improvement (see Table 7).

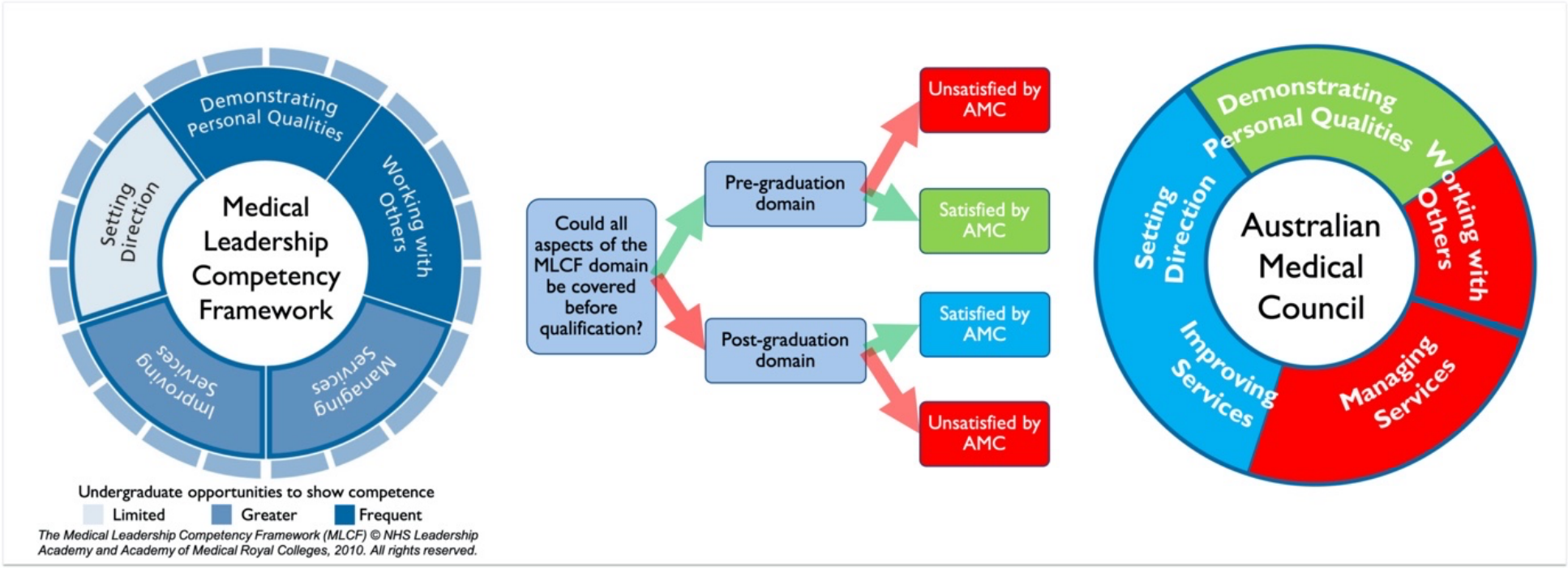


Figure 3: General Needs Analysis Process.
In order to identify curriculum gaps, the authors first assessed the achievability of Medical Leadership Competency Framework (MLCF) domains at an undergraduate level. We then considered whether the Australian Medical Council (AMC) Outcomes for Graduates would satisfy the requirements of each MLCF domain. Where MLCF domains were judged unachievable at undergraduate level, we considered whether foundations were sufficiently laid for future development. 5/8 (63%) of undergraduate MLCF domains and 7/12 (58%) of postgraduate domains were judged to be satisfied by AMC Outcomes for Graduates.

MLCF Domain	MLCF Elements	AMC Outcomes for Graduates
4.1 Ensuring patient safety Assessing and managing the risk to patients associated with service developments, balancing economic considerations with the need for patient safety	1. Identify and quantify the risk to patients using information from a range of sources 2. Use evidence, both positive and negative, to identify options 3. Use systematic ways of assessing and minimising risk 4. Monitor the effects and outcomes of change [audit]	2.14 Place **the needs and safety of patients at the centre of the care process**. Demonstrate safety skills including infection control, graded assertiveness, adverse event reporting and effective clinical handover. **3.6 Describe a systems approach to improving the quality and safety of health care.** 2.5 Select and justify common investigations, with regard to the pathological basis of disease, utility, **safety and cost effectiveness**, and interpret their results. 2.11 Prescribe medications **safely**, effectively and **economically** using objective evidence. Safely administer other therapeutic agents including fluid, electrolytes, blood products and selected inhalational agents.
4.2 Critically evaluating Being able to think analytically, conceptually and to identify where services can be improved, working individually or as part of a team	1. Obtain and act on patient, carer and user feedback and experiences 2. Assess and analyse processes using up-to-date improvement methodologies 3. Identify healthcare improvements and create solutions through collaborative working 4. Appraise options, and plan and take action to implement and evaluate improvements	**3.6 Describe a systems approach to improving the quality and safety of health care.** **1.4 Access, critically appraise, interpret and apply evidence from the medical and scientific literature.**
4.3 Encouraging improvement and innovation Creating a climate of continuous service improvement	1. Question the status quo 2. Act as a positive role model for innovation 3. Encourage dialogue and debate with a wide range of people 4. Develop creative solutions to transform services and care	**3.1 Accept responsibility to protect and advance the health and wellbeing of individuals, communities and populations.** **3.6 Describe a systems approach to improving the quality and safety of health care.** 1.6 Demonstrate a commitment to excellence, evidence based practice and the **generation of new scientific knowledge.** 4.9 **Self-evaluate** their own professional practice; **demonstrate lifelong learning behaviours** and fundamental skills in **educating colleague**s. **Recognise the limits of their own expertise** and involve other professionals as needed to contribute to patient care.
4.4 Facilitating transformation Actively contributing to change processes that lead to improving healthcare	1. Model the change expected 2. Articulate the need for change and its impact on people and services 3. Promote changes leading to systems redesign 4. Motivate and focus a group to accomplish change	**1.6 Demonstrate a commitment to excellence, evidence based practice and the generation of new scientific knowledge.** **3.6 Describe a systems approach to improving the quality and safety of health care.** **3.3 Communicate effectively in wider roles including health advocacy, teaching, assessing and appraising.**

Satisfied Undergraduate		*Satisfied Postgraduate*		*Not feasible at Undergraduate Level*		*Not satisfied*	

Table 5: Example of mapping and matching results.
Only elements judged to be feasible for undergraduates were considered when determining whether a standard was satisfied. The aspects of AMC outcomes that were determined to be most pertinent are in bold text. See supplementary material for full table.

MLCF Domain and description	MLCF Domain Elements
1.2 Managing yourself Organising and managing themselves while taking account of the needs and priorities of others	1. Manage the impact of their emotions on their behaviour with consideration of the impact on others 2. Are reliable in meeting their responsibilities and commitments to consistently high standards 3. Ensure that their plans and actions are flexible, and take account of the needs and work patterns of others 4. Plan their workload and activities to fulfil work requirements and commitments, without compromising their own health
3.1 Planning Actively contributing to plans to achieve service goals	2. Gather feedback from patients, service users and colleagues to help develop plans 4. Appraise options in terms of benefits and risks
4.1 Ensuring patient safety Assessing and managing the risk to patients associated with service developments, balancing economic considerations with the need for patient safety	1. Identify and quantify the risk to patients using information from a range of sources 2. Use evidence, both positive and negative, to identify options 4. Monitor the effects and outcomes of change
4.2 Critically evaluating Being able to think analytically, conceptually and to identify where services can be improved, working individually or as part of a team	1. Obtain and act on patient, carer and user feedback and experiences 2. Assess and analyse processes using up-to-date improvement methodologies 3. Identify healthcare improvements and create solutions through collaborative working 4. Appraise options, and plan and take action to implement and evaluate improvements
4.4 Facilitating transformation Actively contributing to change processes that lead to improving healthcare	4. Motivate and focus a group to accomplish change
5.1 Identifying the contexts for change Being aware of the range of factors to be taken into account	1. Demonstrate awareness of the political, social, technical, economic, organisational and professional environment 2. Understand and interpret relevant legislation and accountability frameworks 3. Anticipate and prepare for the future by scanning for ideas, best practice and emerging trends that will have an impact on health outcomes 4. Develop and communicate aspirations
5.2 Applying knowledge and evidence Gathering information to produce an evidence-based challenge to systems and processes in order to identify opportunities for service improvements	1. Use appropriate methods to gather data and information 2. Carry out analysis against an evidence-based criteria set
5.4 Evaluating impact Measuring and evaluating outcomes, taking corrective action where necessary and by being held account for their decisions	1. Test and evaluate new service options 3. Overcome barriers to implementation

Table 6: MLCF elements deemed suitable for learning before graduation that are missing from AMC Outcomes for Graduates.
For full overview, including elements satisfied, please see Table 16 in Appendix One – Supplementary data table from Chapter Four

Theme Missing	Example MCLF Element
Audit and quality improvement	"Use appropriate methods to gather data and information"; "Test and evaluate new service options"
Team dynamics and working styles	"Take account of the needs and work patterns of others"; "Provide guidance and direction for others using the skills of team members effectively"
Planning and organising	"Plan their workload and activities to fulfil work requirements and commitments"; "Ensure that their plans and actions are flexible"
Understanding the context for change	"Understand and interpret relevant legislation and accountability frameworks"

Table 7: Key MLCF themes missing from AMC Outcomes for Graduates that are suitable for learning during medical school

Discussion

Using the MLCF as a framework for the essential components of medical leadership, we found that many MLCF components are already satisfied within the Australian Medical Council outcomes for graduates. That is, by meeting the AMC outcomes, a student would also meet many components of the MLCF. We also found that there were MLCF domains and elements that could be learnt during medical school but were not yet met sufficiently by AMC outcomes. We have listed these missing MLCF domains and elements in Table 6 and have highlighted them in full in supplementary material (see Appendix). In particular, we noted that quality improvement and audit are missing from the AMC Outcomes for Graduates yet are an essential part of professional practice and aspects could reasonably be learnt during medical school. There were also gaps relating to team dynamics, planning and organising, and understanding change.

The MLCF explicitly contains domains that are beyond the reach of pre-qualification curricula. The original MLCF publication (NHS Institute for Innovation and Improvement, 2010) does not however highlight which specific elements are unlikely to be fully met at undergraduate level. The intention at this level is that undergraduates should have "developed the underpinning knowledge and skills as a foundation for future competence in this area" (NHS Institute for Innovation and Improvement, 2010). A working group in

the United Kingdom has recently highlighted elements of the MLCF curriculum which are attainable in pre-qualification medical curricula. Their report was consistent with our separation of MLCF domains into pre and post-qualification at MLCF domains 1 and 2, while our group judged a greater number of outcomes to be achievable at undergraduate level at MLCF domains 3, 4 and 5 (Peake et al., 2018) (see Appendix One – Supplementary data table from Chapter for our full list of undergraduate and postgraduate outcomes). This could perhaps reflect different programme lengths in New Zealand and the United Kingdom.

Consistent with our findings, the United Kingdom Medical Leadership and Management report highlights the importance of leadership and management skills for newly qualified doctors, in ever changing health systems (Peake et al., 2018). While our study is set in an Australasian context, taken together with the United Kingdom report our findings are generalisable. Our study and the United Kingdom report both present findings supporting leadership learning in all pre-qualification medical curricula so that newly graduated doctors are equipped to work in changing health systems, irrespective of where they graduate. Much of leadership learning is generic and therefore existing frameworks such as the MLCF could be adapted for country specific contexts with relative ease.

Domains requiring postgraduate experience could benefit from more clearly laid and explicit foundations to help students contextualise leadership within their future practice. For example, students may not be able to “overcome barriers to implementation” (MLCF Element 5.4.3, see Table 2). It would however be helpful to introduce implementation of change, so that they may be able to see specific barriers, begin to understand why they persist, and begin to consider how to address them.

While the MLCF did not explicitly highlight which elements could be achieved as undergraduates, it remained the closest to an internationally accepted framework for

medical leadership. When matching to the AMC Outcomes for Graduates, we were forced to interpret standards that were often ambiguous in their intention. In order to reduce bias from this interpretation, we conducted our mapping and matching in parallel, and referred to each other frequently to obtain consensus, as well as referring to MLCF authors. Our review only considered Australian and New Zealand curriculum standards. Internationally, there are a number of similar medical curricula which may be able to draw comparisons, particularly the United Kingdom (General Medical Council, 2018). Given the variation in medical school curricula, we did not specifically address ways in which missing elements could be introduced, nor did we address the outcomes that should be assessed as part of programmes. As assessment often drives learning (Wormald et al., 2009), it will be important for medical schools to develop a strategy for assessment in their local context.

A range of definitions and frameworks for medical leadership have been developed in recent years. Since the MLCF was developed in 2010, medical leadership curriculum frameworks have been produced by the General Medical Council (General Medical Council, 2012), National Health Service (NHS) National Leadership Council (National Health Service National Leadership Council, 2011), NHS Leadership Academy (National Health Service Leadership Academy, 2013), Faculty of Medical Leadership and Management (FMLM) (Faculty of Medical Leadership and Management, 2016), RACMA (Royal Australasian College of Medical Administrators, 2011) and the Canadian College of Health Leaders (Canadian College of Health Leaders, 2010), as well as by many other medical councils and colleges. Yet, our analysis shows that 38% of undergraduate domains and 42% of postgraduate domains from the 2010 MLCF were still not met by the current case example of the AMC Outcomes for Graduates. It appears that there has perhaps been more progress with development of medical leadership frameworks than there has been with development of medical leadership curricula in medical schools.

In conclusion, while current Australian Medical Council Outcomes for Graduates meet many elements of medical leadership as set out in the MLCF, this needs analysis indicates that they do not sufficiently cover many other elements. Leadership domains already satisfied by the AMC Outcomes could benefit from being highlighted within AMC curricula as being relevant to leadership, in order for students to begin to understand where leadership skills and knowledge fit into their future practice. The MLCF guidance for undergraduates suggests a number of ways to introduce missing elements into already cramped programmes, as does the recent United Kingdom Faculty of Medical Leadership and Management publication by Peake et al. (Peake et al., 2018; Spurgeon & Down, 2010).

This needs analysis has implications both for the research in this thesis, as well as for medical schools directly.

Implications for medical schools

We suggest the following approach could be taken by medical schools:

1. Agree that medical leadership is an important domain of learning for pre-qualification medical students;

2. Clearly define the outcomes at each level of learning that is, pre-qualification, early post qualification and postgraduate (this will be context specific). Use the MLCF as a common document to agree graduate outcomes, as it is the most well-established and robust framework for medical leadership;
3. Establish a local baseline. As well as comparison to MLCF missing elements, focus groups with local medical students, teaching faculty and teaching hospitals could offer important insights (Rohwer et al., 2014; Thomas et al., 2016);
4. Prospectively plan evaluation of the curriculum modification;

5. Modify approach to medical leadership development. Medical schools are free to exceed medical council requirements. Given variation in contexts and approaches in each school, it may be most effective for individual programmes to develop their own approach that fulfils the modified accrediting body and MLCF standards;
6. Evaluate the outcomes of modifying the curriculum;
7. Modify the standards and implementation based on the evaluation outcome data;
8. Work with accrediting bodies such as the AMC to implement modified graduate outcomes;

Further work in Australia and New Zealand could look specifically at the postgraduate curricula (Medical Council of New Zealand, 2014), to assess the degree to which post-qualification elements are addressed after students graduate and become doctors. This would help to ensure continuity with undergraduate programmes and to establish a shared language around leadership.

Chapter Conclusion

The needs analysis presented in this chapter showed that the current medical curriculum standards in Australia and New Zealand fail to cover medical leadership development competencies as outlined in the MLCF. Similar gaps were mirrored in a United Kingdom-based project as described above, showing that this is not purely an Australasian medical school gap in leadership development. This means that unless medical schools in Australia, New Zealand and the United Kingdom aim beyond these standards, they will not necessarily equip their graduates with necessary foundations for leadership aspects of their future roles as informal or formal medical leaders.

There is marked variation between individual medical schools however, and many may choose to update their curricula well in advance of changes to medical council standards for graduates. To explore the range of ways in which medical schools have already integrated and evaluated leadership development programmes for their students, I conducted the systematic review presented in Chapter Five.

As I have explained in earlier chapters, having established that there was a gap in medical leadership curriculum standards in New Zealand and Australia, I shifted my focus to research medical leadership development for medical students internationally with the systematic review I have presented in Chapter Five. In later chapters, with a view to producing research with high potential for impact, I will shift my research focus further, to focus on medical leadership development for doctors in the NHS (Chapters Six, Seven, Eight and Nine) and internationally (Chapter Ten)

Chapter Seven – Retrospective Self-Assessments

Chapter Introduction

In Chapter Six, I explored the development of the Revised Primary Colours Questionnaire and piloted its use in the Emerging Leaders Programme Cohort One (ELP1). In Chapter Seven, I will introduce the use of retrospective self-assessments to adjust for response-shift bias, and I will pilot the use of retrospective self-assessments in the Emerging Leaders Programme Cohort Two (ELP2).

As explored in Chapter Six, there was a general decrease in the Original Primary Colours Questionnaire results from the early-programme assessment to the post-programme assessment. This was surprising given extremely positive feedback from participants and an independent evaluation of team projects on the FLP which showed substantial return on investment in financial terms (Canter, 2015, 2016). Qualitative data from focus groups, interviews, observations and free-text comments showed that participants reacted positively to the programme, considered their confidence and leadership skills and knowledge to have improved, had applied learning into their day to day behaviour and achieved organisational results (Canter, 2015, 2016).

In focus groups and informal interactions, participants suggested that when they had first rated themselves (second FLP 2016-2017 session) they might have overestimated their competence. They suggested that as they learned through the programme, the internal scales they measured themselves against changed. They proposed this could have explained why their post-programme ratings tended to be lower than their original pre-programme ratings when they were asked to self-assess their competence at the conclusion of the programme.

Such an effect has been termed “response-shift bias” and is the topic of Chapter Seven.

Response-shift bias

Any test that compares a pre- and post-intervention result relies on the presence of a stable test instrument. In physics, the stability of mass, time and distance measurements are quality assured by comparison to a stable, agreed, set of universal references. Standards and universal references are employed in many laboratories concerned with measurement. Self-assessments, however, often do not have a stable reference point. Researchers using pre- and post-programme self-assessments must assume that subjects have a stable internal standard reference point for their responses that will not change between the self-assessments. If subjects undergo an intervention that influences their internal reference points there will be a resultant shift in their self-assessments: “Response-Shift Bias” (Howard et al., 1979).

Response shift bias can be related to the experience of someone being outside for too long on a cold day. When they return inside and put their hands into warm water, the water feels boiling hot, even though it only felt warm before they went into the cold outside. Their internal reference point has shifted to match the cold temperature outside, meaning their assessment of temperature now (“boiling”) is inconsistent with their previous assessment (“warm”).

Retrospective ‘then-test’

The retrospective pre-test or ‘then-test’ was first introduced in the mid twentieth century as a means of obtaining a pre-intervention measurement when traditional pre-tests were not possible (Campbell & Stanley, 1963; Nimon & Allen, 2007). Retrospective ‘then-tests’ were later proposed as a means to detect and adjust for response-shift bias

(Howard et al., 1979). ‘Then-tests’ have since became popular as an evaluation methodology in parenting (Geldhof et al., 2018; Pratt, 2002), education (Allen & Nimon, 2007; Bhanji et al., 2012; Lam & Bengo, 2003; Nakonezny & Rodgers, 2005; Skeff et al., 1992) and also in leadership development (for example, Goedhart & Hoogstraten, 1992; Kim Nimon, Zigarmi, & Allen, 2011; Rohs, 2002; Taylor, Russ-Eft, & Taylor, 2009).

While several studies have used ‘then-tests’ in evaluation of medical leadership development programmes (Drennan, 2012; Fernandez et al., 2016; Frugé et al., 2010; Gilfoyle et al., 2007; Haftel et al., 2018; Hopkins et al., 2018; Osborn & DeWitt, 2004; Pettit et al., 2011), these studies have been criticised for lack of validation or triangulation with objective findings (Hill, 2020). Indeed, many studies which compare pre-post and then-post changes conclude that ‘then-post’ evaluations show more statistically significant changes, therefore they are more accurate (for example, Drennan, 2012; Drennan & Hyde, 2008; Rohs, 2002). Little consideration is given to the possibility that ‘then-post’ evaluations might inflate intervention impact.

Objectives

The purpose of this study was to evaluate the use of the retrospective ‘then-test’ (referred to from this point without quotation-marks) as an evaluation method for medical leadership development. My aims were specifically to:

1. Compare changes in self-assessed leadership questionnaires measured by traditional pre-post evaluation with changes measured by retrospective then-post evaluations; and
2. Evaluate concordance with a parallel mixed-methods evaluation to allow for consideration of possible inflation of impact when using the then-test.

Case study

For this chapter, the evaluation methods described were applied in ELP2, an 8-month long leadership development programme for 17 multidisciplinary health professionals. Further information on ELP2 is in Appendix Six – Case study descriptions.

Evaluation methods

The Revised Primary Colours Questionnaire (see Chapter Six) and an adapted form of the Medical Leadership Competency Framework (MLCF) Self-Assessment Tool (NHS Leadership Academy, 2012) were both applied as pre-test, post-test and then-test self-assessments as shown in Figure 4 and Table 8.

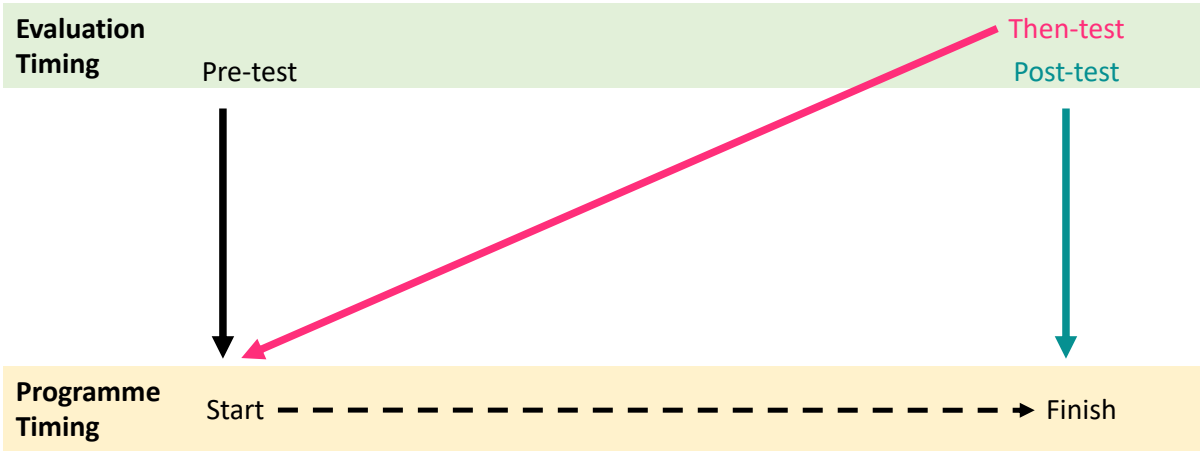


Figure 4: Explanation of retrospective pre-test (then-test).
Pre-test refers to traditional prospective pre-test. Then-test refers to retrospective pre-test. The pre-test and then -test are compared with the post-test

I reflect on how my own values and principles influence my behaviour and impact on others.

	1	2	3	4	5	6	7
BEFORE the programme	○	○	○	○	○	○	○
AFTER the programme	○	○	○	○	○	○	○

Figure 5: An example of a post-course question from the Medical Leadership Competency Framework Self-Assessment incorporating both a post-test ("AFTER") and a then-test ("BEFORE"). 1=strongly disagree, 7=strongly agree.

These questionnaires were administered in parallel pre programme (pre-test) and post-programme (both post-test and then-test). In the post-programme questionnaires,

each question was presented as a pair of responses, one for post-programme and one for pre-programme (see Figure 5).

Programme impact was concurrently evaluated using free-text questions, interviews, observations and evaluation of participant project work (see Table 8). This approach was based on Kirkpatrick's Framework for training programmes (Steinert et al., 2012; Yardley & Dornan, 2012). To align with the questionnaires being piloted, the additional data collected focused on learning (aligned with the Primary Colours Questionnaire) and behaviour (aligned with the MLCF). Outcomes were also assessed at Level Four (Results) as a matter of course. Basic demographic questions were included in questionnaires, along with open questions relating to programme engagement and experience. Completion of the post-programme evaluations was required as part of the programme.

Data source	Data Type	Timing in programme			Kirkpatrick Levels			
		Pre	During	Post	Reaction	Learning	Behaviour	Results
Revised Primary Colours Questionnaire	Likert scale	✓	✗	✓(post, then)	✗	✓	✗	✗
MLCF Self-Assessment	Likert scale	✓	✗	✓(post, then)	✗	✗	✓	✗
Supplementary Questionnaire*	Mixed	✗	✗	✓(post, change)	✓	✓	✓	✓
Free-text questions	Free-text	✗	✓	✓	✓	✓	✓	✓
Observations	Observation	✗	✓	✓	✗	✗	✓	✓
Project Outcomes	Mixed	✗	✗	✓	✗	✗	✓	✓

Table 8: Additional methods used to evaluate Emerging Leaders Programme Cohort Two. MLCF, Medical Leadership Competency Framework. *See Chapter Eight for more information on the supplementary questionnaire.

Ethical approval

Participant responses were anonymised and linked by the use of pre-assigned participant codes. The University of Oxford Clinical Trials and Research Governance (CTRG) was consulted and ethics exemption was granted as an evaluation of service delivery.

Analysis

For the MLCF Self-Assessment tool, the eight elements of each of the seven domains were summed to produce a combined score for each domain between 8 and 56. These MLCF domain scores were then scaled down to a score of 1-7 for ease of interpretation.

Statistical analysis

Significance was calculated using Wilcoxon Signed-Rank Test, an appropriate test for non-parametric paired measurements. Alpha was defined as p=0.05 and corrected for multiple comparisons using the Bonferroni correction in each set of tests (Primary Colours Domains, Primary Colours Tasks, MLCF Domains). See Table 9 for adjusted p-values.

Significance tests were applied separately to pre-post test pairs and then-post test pairs from each questionnaire.

Qualitative analysis

Thematic analysis (Braun & Clarke, 2006) was applied to free text comments using Kirkpatrick Levels Two (Learning) and Three (Behaviour) as sensitising concepts.

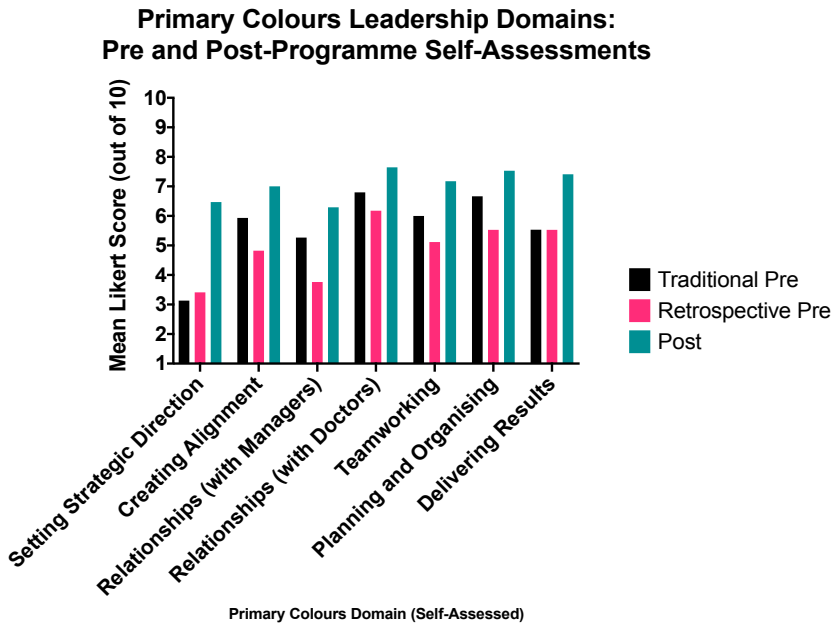
Results

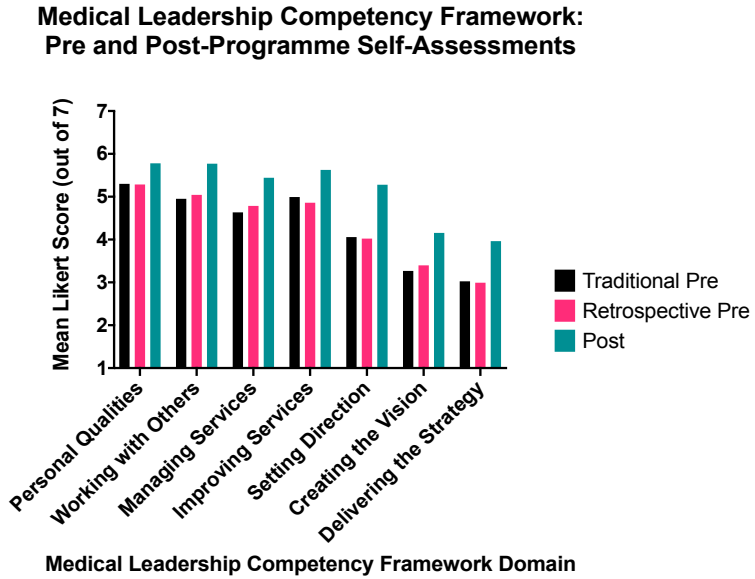
When compared to post-test self-assessments, retrospective “then-test” self-assessments demonstrated a greater number of statistically significant changes than traditional prospective pre-test self-assessments.

Specifically, changes reaching statistical significance increased from 4/12 items pre-post to 11/12 items then-post (Revised Primary Colours) and from 3/7 pre-post to 7/7 then-post (MLCF) (Table 9 and Figure 6).

	Post-Programme	Pre-Programme (Prospective)		Pre-Programme (Retrospective)	
	Mean Score	Mean Increase	P value	Mean Increase	P value
Primary Colours Domains (rated 1-10)					
Setting Strategic Direction	6.47	3.20	0.0006	3.06	0.0004
Creating Alignment	7	1.00	0.0311	2.18	0.0031
Relationships (with Managers)	6.29	0.93	0.0558	2.53	0.0015
Relationships (with Doctors)	7.65	0.67	0.1122	1.47	0.0074
Teamworking	7.18	0.93	0.1191	2.06	0.0009
Planning and Organising	7.53	0.73	0.0873	2.00	0.0013
Delivering Results	7.41	1.67	0.0030	1.88	0.0013
Primary Colours Tasks (rated 1-10)					
Inspire	7.12	1.33	0.0052	2.00	0.0004
Focus	6.94	1.47	0.0062	1.59	0.0028
Enable	6.76	1.00	0.0169	1.82	0.0008
Reinforce	6.71	1.27	0.0239	1.59	0.0010
Learn	8.06	1.33	0.0243	1.65	0.0018
Medical Leadership Competency Framework Domains (rated 1-7)					
Personal Qualities	5.78	0.43	0.0108	0.49	0.0021
Working with Others	5.77	0.75	0.0037	0.73	0.0006
Managing Services	5.44	0.77	0.0058	0.65	0.0004
Improving Services	5.63	0.63	0.0100	0.76	0.0006
Setting Direction	5.28	1.09	0.0032	1.26	0.0004
Creating the Vision	4.15	0.78	0.0105	0.76	0.0004
Delivering the Strategy	3.96	0.77	0.0355	0.97	0.0004

Table 9: Significance tests for Primary Colours Domains, Primary Colours Tasks, and Medical Leadership Competency Framework Domains (ELP2). Statistical significance was tested using the Wilcoxon signed-rank test for paired pre-post (prospective) and then-post (retrospective) self-assessments. The conservative Bonferroni correction was applied to reduce type I error (false-positives) for each set of tests, with number of tests n=7 for Primary Colours Domains Tests (adjusted p=0.0071), n=5 for Primary Colours Tasks Tests (adjusted p=0.01), and n=7 for MLCF tests (adjusted p=0.0071) Significant changes are highlighted in green. Note the greater proportion of items that are statistically significant when using retrospective then-post test pairs.





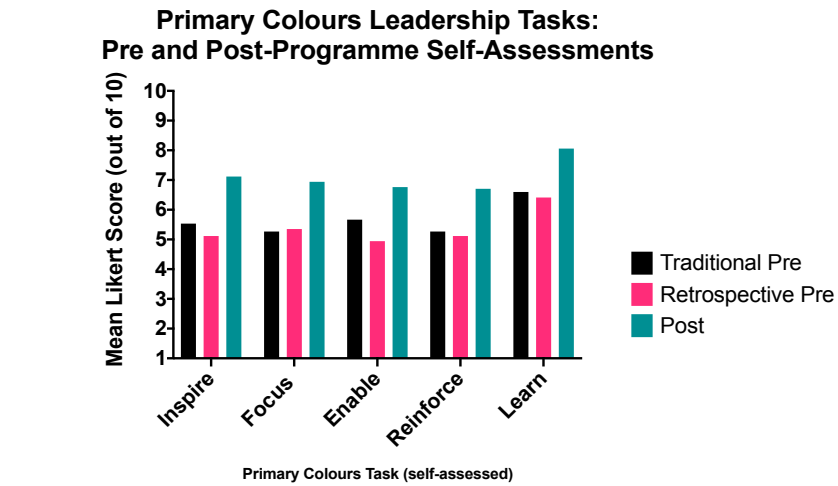


Figure 6: Traditional Pre-, Retrospective Then-, and Post- test self-assessments for the Primary Colours Questionnaire and the Medical Leadership Competency Framework.
Primary Colours Domains and Primary Colours Tasks were anchored at 1=poor and 10=excellent; Medical Leadership Competency Framework Domains were anchored at 1=poor and 7=excellent. Note that there tends to be a greater increase between retrospective pre (red) and post (cyan), and note that there is less difference between traditional and retrospective pre in the MLCF questionnaire (top right panel). For more detail please see Table 9.

Additional measures of impact

Kirkpatrick Level Two: learning

While the Revised Primary Colours Questionnaire more specifically measures self-assessed knowledge and skills, Kirkpatrick Level Two (learning) also includes changes to attitudes.

In free text comments relating to attitude, ELP2 participants particularly cited improved confidence, feeling “empowered” and changed perspective on leadership.

In the supplementary questionnaire (see Chapter Eight), three quarters (13/17) of ELP2 participants reported that their confidence had increased as a result of the programme, and four-fifths (14/17) reported that their motivation had increased as a result of the programme. More than half of ELP2 participants (11/17) reported no change in their job satisfaction or resilience. One participant reported that their job satisfaction had decreased as a result of the programme (see Chapter Eight).

In free-text comments relating to skills and knowledge, ELP2 participants reported improved knowledge about leadership and management and engagement with stakeholders, as well as improved self-awareness and understanding of health system structure: “It was invaluable in providing insight into the structural processes involved in effecting change at individual, Trust and even National level.”

Kirkpatrick Level Three: behavioural changes

ELP2 Participants reported spending an additional 60-120 minutes per week (median) outside their clinical responsibilities working on their ELP projects. Most ELP2

participants (15/17) reported having been involved in or having started clinical improvement projects other than their ELP project during the programme.

Free-text comments reported that ELP2 participants were using tools learned on the ELP for their other projects. Several free-text comments also mentioned that the programme had afforded participants “good networking opportunities”, and participants reported using these connections in their day to day work: “I have learnt tools for other ongoing/background projects, and some things I wish I had known before I started!”

Two ELP2 graduates proceeded to organise and deliver workshops for the Emerging Leaders Programme Cohort Three (ELP3), and three further graduates contributed as individual leadership coaches (29%). Given that none of these graduates had led leadership development programmes or had coached before they graduated from the ELP, this constitutes a notable change in behaviour.

Three programme participants were promoted in their professional role during the programme, including both of the ELP2 nurse participants (promoted to matron) and one of the senior medical trainees (promoted to consultant).

Kirkpatrick Level Four: organisational results

One participant successfully advocated for Health Education England Thames Valley (HEETV) to fund the next cohort of the ELP (ELP3), and for all future HEETV Deanery Quality Improvement Fellows to attend as a required component of their fellowship.

One participant reported “I found it difficult to implement change in a department/area of medicine that I had limited knowledge/experience”. Projects had variable success during the programme (see Table 10), with one project being formally approved for scaling and funding throughout the hospital, and one team not managing to achieve their goal at all.

Project	Impact
Peer Support Groups for Staff	Well-received by attendees, leading to hospital-wide approval for implementation. Successfully gained funding for the next 2 years.
Improving Peri-Natal Wound Care	Evaluated provider and patient confidence managing perinatal perineal wounds and produced information sheet for new mothers regarding appropriate wound care.
Reducing Peri-operative Respiratory Arrests after Spinal Opiates	Conducted audit of practice, established local guideline for dosage of spinal opiates and presented this at a departmental meeting. Electronic alerts implemented on the prescribing system for when dosage guidelines exceeded.
Support Groups for Vulnerable Pregnant Women	Interest in support groups was established, as well as contact with possible sponsors. No support group was successfully implemented.
Homeless Care Pathways	Emergency staff improved in their confidence managing the discharge of homeless patients, and the number of representations decreased.

Table 10: ELP2 Project outcomes.
Impact has been summarised from project presentations and posters.

Discussion

This study showed a considerable increase in the proportion of items showing statistically significant changes using a retrospective then-test compared to a traditional pre-test. This finding held across both questionnaires used: the Primary Colours (6/7 vs 2/7 domains and 5/5 vs 2/5 tasks); and the MLCF Domains (7/7 vs 3/7 domains).

The mixed-methods evaluation conducted alongside pilot of then-tests in the Primary Colours and MLCF Self-Assessment Questionnaire showed positive outcomes in free-text comments, project outcomes, participant promotions, and participants supporting the following cohort (ELP3) in faculty roles, and in one participant successfully securing funding and support for future cohorts of the ELP. Taken together, the mixed-methods data suggest that the programme was broadly successful without specifically offering objective corollaries for the questionnaire items.

There are a range of proposed reasons why then-tests might show a greater number of statistically significant changes. Early work by Howard and colleagues showed that

retrospective then-tests were more closely aligned with changes in objective performance than traditional pre-tests (Howard et al., 1979). This finding has been replicated in a range of settings (Geldhof et al., 2018; Hill, 2020; Pratt et al., 2000; Taylor et al., 2009). One reason that then-tests show a greater number of statistically significant results could therefore be that they are more accurate in portraying changes across a programme than traditional prospective pre-tests.

Many prominent commentators have, however, urged caution in accepting then-tests as being more objective and accurate indicators for change than traditional pre-tests, without careful validation against objective criteria.

While our mixed methods evaluation data generally indicate a positive impact of the programme, our methods did not include objective measures of the specific constructs measured by the Primary Colours or MLCF instruments. As explored in Chapters Two and Three, there is a lack of established objective evaluation tools that could have been employed.

While there are indications that the programme achieved positive results, it is not possible to infer the comparative accuracy of the then-tests and the pre-tests through comparison with an objective set of criteria. This does not mean that it is not possible to find objective measures which could be correlated with questionnaire items in the future, and this could be an important area for future research if the Primary Colours and MLCF Questionnaires continue to be used.

There are further risks of simply accepting increased detection of changes as being more accurate. While then-tests may control for response-shift bias, when used as a replacement for traditional pre-tests there is a risk of introducing other biases, as outlined below. To improve the detection of these biases, most researchers therefore recommend that then-tests be applied only in addition to traditional pre-tests, not as a replacement

(Campbell & Stanley, 1963; Geldhof et al., 2018; Hill, 2020; Hill & Betz, 2005; Norman, 2003; Taylor et al., 2009).

Participants are prone to holding an implicit theory of change. Individuals who are aware that they have been subject to an intervention will tend to document change using a then-test, even when they are in a placebo group where the intervention was unrelated to the self-assessment. This relates closely to the classic Hawthorne effect (The Hawthorne Effect in Educational Research, 2015; Roethlisberger et al., 1939). Because the participants on the ELP were aware they had completed a leadership programme, they would likely have been primed to expect an improvement in their leadership competency.

Where participants have invested time and effort in an intervention, they may be prone to error-justification bias (Hill & Betz, 2005). Because ELP participants had invested a considerable amount of their free time working on projects and attending workshops, they would have been prone to justifying that investment by considering themselves to have improved in their leadership competencies.

Where participants perceive an improvement to be sociably desirable, there is a tendency for them to inflate improvements, both to please assessors and to enhance their self-opinion. This has been termed social desirability bias or “impression management” (Pearson et al., 1992). Because the faculty of the ELP were clearly invested in helping participants to develop as leaders, there may have been social pressure for participants to deflate their then-test scores to inflate the measured change in leadership competency.

People also tend to be more critical of their past than their present self, which could result in general deflation of then-tests, which would inflate change observed using a then-test (Wilson & Ross, 2001).

It is interesting to note that while then-tests showed more consistent statistical significance that pre-tests in the MLCF self-assessments, the absolute change between the pre-test and the then-test was minimal, in some cases showing higher then-test scores than pre-test scores (see Figure 6). Two factors are particularly of note here.

Firstly, while the Primary Colours instrument consists of seven single domain items and five single task items, where participants are asked to rate themselves on the construct (for example, planning), the MLCF domains consist of eight questions per domain with a focus on behaviours rather than constructs (for example, “I provide others with clear purpose and direction”; “I assess the available options in terms of benefits and risks”). There are indications that behavioural items are less susceptible to response-shift bias than items relating to subjective ratings or attitudes (Bornstein et al., 2020; Schwartz & Rapkin, 2004). This could indicate that the increased number of statistically significant items on the MLCF Self-Assessment using then-tests indicates not reduced response-shift bias, but increased bias from implicit theory of change, social desirability, effort justification. In order to make a valid judgement as to whether response-shift was adjusted for or another bias accentuated, we would need to develop a viable objective measure.

Secondly, the MLCF component of the questionnaire contained seven sets of eight questions, in both then-test and post-test form, for a total of 112 answers required across the then- and post- test responses to each of the 56 questions. When respondents are faced with more difficult and numerous tasks, there is greater tendency towards satisficing. The Wilcoxon signed-rank test relies not on the magnitude, but the consistency of change, by assigning positive rank (increase) or negative rank (decrease) to each pair and calculating the probability that this number of positive ranks could be obtained by random chance. Therefore, if participants resorted to satisficing and merely ranked each of their then-tests at one point less than their post-test (out of seven), there would be greater likelihood of statistical significance, even with small magnitudes of

change. This possibility could perhaps best be tested by cognitive interviewing (Willis, 2005).

Additionally, while both surveys applied in this study were designed according to Howard's initial recommendation that then-tests are conducted side-by-side with post-tests (Howard et al., 1979), there is recent evidence that suggests separation of then-test and post-test questionnaires into two forms, filling each out without reference to the other, is less prone to the aforementioned biases (effort-justification, social desirability, implicit theories of change and critical self-reflection, and satisficing) (Lam & Bengo, 2003; Nimon et al., 2011; Schwartz & Sprangers, 2010).

There are several practical implications of these results:

- Where it is only possible to administer a post programme evaluation, the inclusion of then-test questions may provide useful additional information, with minimal additional effort (Campbell & Stanley, 1963; Geldhof et al., 2018; Hill & Betz, 2005);
- When designing the then-test for post-programme administration, recent research indicates that separate then-test and post-test forms are more reliable than adjacent then- and post- items (Lam & Bengo, 2003; Nimon et al., 2011; Schwartz & Sprangers, 2010);
- Where possible a traditional pre-test should also be administered before the start of the programme, as additional biases may not be elucidated through the use of then-tests alone.

Chapter Conclusion

In Chapter Six, the possibility of response-shift bias affecting the results of a pre-and post-programme application of the Revised Primary Colours Questionnaire was introduced. In Chapter Seven, retrospective then-tests were added to the post-

programme Primary Colours Questionnaire, and significance tests supported the possibility of response-shift bias, with important caveats regarding other biases as above.

In addition to the Revised Primary Colours Questionnaire, a second questionnaire (the MLCF self-assessment) was adapted to include a then-test in the post-programme assessment to investigate the possibility of a response-shift bias in an additional questionnaires (that is, not confined to the specific Revised Primary Colours Questionnaire). Again, the significance tests presented in this chapter supported the possibility of a response-shift bias in this second questionnaire also.

From the analysis presented in tables and figures of the pre-, then- and post-test self-assessments, it appears that the MLCF questionnaire responses are less prone to changing internal reference points (response-shift) than the Revised Primary Colours Questionnaire. Two possible causes for increased vulnerability in the Primary Colours Questionnaire were suggested in the discussion: single-construct questions, and construct rather than behavioural questions.

In order to determine whether the increase in the number of statistically significant changes using the then-test reflects a reduction in response-shift bias or an introduction of additional biases as suggested in the discussion above, it would be helpful to repeat this study with objective indicators to match some or all of the items in the two questionnaires.

Given that such objective indicators have proven difficult to find, introducing control questions that are not expected to change could also help quantify the degree of change-inflating biases (such as, trivially, “how happy were you with your high-school education?”). If a change-inflating bias is causing the increased changes, then these carefully chosen unrelated control questions would be expected to increase as well as the questions directly relating to programme outcomes.

Broadly, this chapter has indicated several potential methodological improvements that could be implemented into leadership development self-assessment questionnaires such as the Revised Primary Colours Questionnaire developed in the Chapter Six, including the addition of retrospective then-tests. For such questionnaires to offer useful information for decision-making, further development is needed. In Chapter Nine, I will re-evaluate the expectations and needs of key decision-makers, with a view to guiding further development of evaluation methods such as the Revised Primary Colours Questionnaire.

First, however, in Chapter Eight I will describe a second questionnaire which I developed alongside the Revised Primary Colours Questionnaire. This second questionnaire was briefly introduced as the “supplementary questionnaire” in this chapter and in Table 8 of Chapter Six.

Chapter Eight – Supplementary Questionnaire Development

Chapter Introduction

Chapter Seven explored response-shift bias in the evaluation of medical leadership development, after the evaluation of the Future Leaders Programme (FLP) (a leadership development programme for recently appointed consultants) indicated that there could have been significant response-shift bias in self-assessed outcome measures. The focus in Chapter Seven was the Revised Primary Colours Questionnaire (domains and tasks) and on the MLCF Self-Assessment.

In the evaluation of the FLP 2016-2017, first introduced in Chapter Six ("Development of the Revised Primary Colours Questionnaire"), participants noted that they had been affected in a range of ways that were not captured in the Original Primary Colours Questionnaire. In particular, FLP 2016-2017 participants noted that their subjective wellbeing and satisfaction in their jobs had increased, and that they had applied their learning into quality improvement projects outside the programme.

In order to explore these suggested outcomes, I formally interviewed one of the FLP 2016-2017 participants who had drawn this unexplored impact to my attention and co-designed a Supplementary Questionnaire which captured data around the themes they suggested.

The aim of this chapter is to describe the development of this Supplementary Questionnaire, designed to measure impacts of a leadership development programme in healthcare that were not captured in the Original or the Revised Primary Colours Questionnaire.

Methods

In order to develop a contextual understanding of a leadership development programme in healthcare and to develop more focused research questions, I obtained verbal consent from the faculty and participants of the FLP 2016-2017 to observe and evaluate this programme. I was introduced to programme participants as a researcher by the programme faculty, who encouraged participants to contribute to our understanding of medical leadership development through informal discussions with me in addition to the formal programme evaluation.

As my research centred on evaluation of a programme funded by public money, I was granted ethics exemption for this research by the Clinical Trials Research Governance Committee (University of Oxford).

Phase One: Scoping

One month before the end of the (12-month) programme, I was approached by a participant (JB) for an informal discussion about programme outcomes that he felt had not been adequately captured in the existing evaluation. He agreed to a formal interview to explore incorporation of these outcomes into our evaluation methods. The 38-minute interview was conducted at JB's workplace, during his lunch break.

I recorded and transcribed the interview verbatim and sent the transcription to JB for checking. I coded the interview transcription in NVivo 12 using a systematic text condensation approach (Malterud, 2012), and identified major and minor themes, as shown in Table 11.

Major Themes	Minor Themes
Impact outside the programme	*Clinical activity*
	Departmental activity
	Personal life
Impact on wellbeing	*Confidence*
	Motivation
	Job Satisfaction
	Resilience

Table 11: Major and minor themes identified in interview with Future Leaders Programme participant

Phase Two: Questionnaire development

After identifying these themes, I developed single-item questions for a supplementary questionnaire, employing three-point Likert Scales alongside free-text questions.

There is added granularity and statistical power from wider Likert scales, as used in the adapted Primary Colours Questionnaire (see Chapter Six for justification of the 10-point scale). I opted to use three-point scales in the Supplementary Questionnaire for consistency with the Original Primary Colours Questionnaire which was used in FLP 2016-2017 (Chapter Six), because the pilot introduction of this questionnaire was for the same FLP 2016-2017 cohort.

In order to improve the design of this early questionnaire, I checked the question wording and content with the author of the Original Primary Colours Questionnaire (RC) as well as with JB. Through a series of iterations, we improved the clarity of question wording. This did not constitute full validation of the questionnaire. The process of checking and improving working did, however, increase the usability and face validity of the questionnaire. The Supplementary Questionnaire as applied in the FLP 2016-2017 is attached to this thesis as Appendix Five – Supplementary Questionnaire.

After piloting the Supplementary Questionnaire with two programmes (FLP 2016-2017 and ELP1), I reduced the number of questions before use in ELP2 and ELP3, as some questions were not as generalisable to other programmes.

Phase Three: Questionnaire administration

I employed the Supplementary Questionnaire in four case studies. Respondent demographics are listed in Table 12.

	Programme			
	FLP	**ELP1**	**ELP2**	**ELP3**
Respondents n (%)	18 (75%)	13 (93%)	17 (100%)	20 (77%)
Discipline (% of respondents)				
Consultants	18 (100%)	-	-	-
Trainees	-	13 (100%)	11 (65%)	13 (65%)
Nurses	-	-	2 (12%)	4 (20%)
Other	-	-	4* (24%)	3** (15%)
Other Data Collected	Free Text 6 Interviews	Free Text	Free Text	Free Text BRS Pre Post

Table 12: Demographics of respondents to the Supplementary Questionnaire. *Two midwives and two physiotherapists. **One occupational therapist, one physiotherapist and one research manager. FLP, Future Leaders Programme. ELP1, Emerging Leaders Programme Cohort 1, etc. BRS, Brief Resilience Scale

All four case study evaluations included free-text questions which were included in the qualitative analysis. Additionally, I conducted interviews with six graduates of the FLP (2016-2017). These interviews lasted 30-40 minutes and were recorded and transcribed verbatim and checked with interviewees. One interview was conducted with a dyad (two individuals at once), and one interview was conducted by phone. The remaining three interviews were conducted face-to-face with single interviewees.

In order to investigate criterion-validity of the resilience item on the supplementary questionnaire, the validated Brief Resilience Scale (Smith et al., 2008) was employed alongside the Supplementary Questionnaire in the pre- and post-programme evaluation for Emerging Leaders Programme Cohort Three (ELP3).

Case studies:

1. Future Leaders Programme (FLP 2016-2017) for newly appointed consultants at OUH;

2. Emerging Leaders Programme cohort 1 (ELP1, 2018) for PGY1-4 doctors in leadership roles at OUH;
3. Emerging Leaders Programme cohort 2 (ELP2, 2019) for any pre-consultant medical trainees in leadership roles at OUH and including a small number of health professionals from non-medical clinical backgrounds. This questionnaire was moved to an online form at the request of participants and was reduced in length as described above;
4. Emerging Leaders Programme cohort 3 (ELP3, 2020) with an expanded proportion of health professionals from non-medical clinical backgrounds.

See Appendix Six – Case study descriptions for more information about the case studies.

Analysis

I analysed quantitative items on the Supplementary Questionnaire using descriptive statistics.

I conducted a thematic analysis (Braun & Clarke, 2006) of free text answers from both the Supplementary Questionnaire and the main end-of-programme evaluations using the themes developed in the initial interviews with JB (Table 11) as sensitising concepts.

The Brief Resilience Scale pre-and post-programme scores were analysed using the Wilcoxon signed-rank test for non-parametric data with an alpha of 0.05.

Ethics

Ethics exemption was granted by the University of Oxford Clinical Trials Research Governance committee (CTRG) for each of the case studies as evaluations of service delivery.

Results

Impact outside main programme

Fifteen of the 18 FLP respondents (82%) had started clinical improvement projects outside their main FLP project, as JB had suggested might be the case. All respondents reported they had applied skills learned on the FLP in areas outside their main project, as shown in Figure 7.

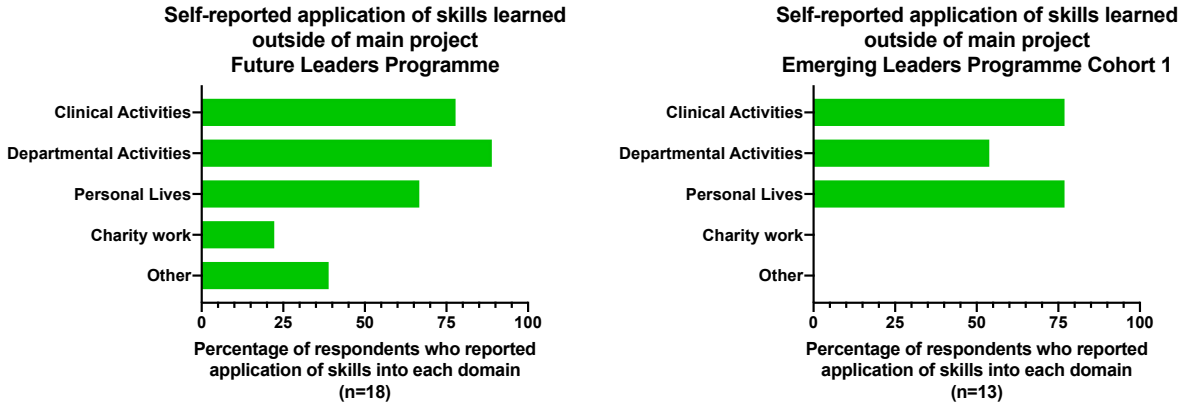


Figure 7: Impact outside the main programme for the Future Leaders Programme (2016-2017, left) and the Emerging Leaders Programme Cohort One (right).
This question was removed from the ELP2 and ELP3 questionnaires, to remove redundancy with a related free-text question.

A considerable majority of FLP 2016-2017 participants reported further application of their learning into departmental activities, their personal lives, charity roles or other positions (sports coaching x2, national roles x2). A slightly smaller majority of respondents from the ELP reported application into departmental activities, as shown in Figure 7.

Impact on wellbeing

All FLP 2016-2017 respondents reported that participating on the programme had increased their motivation, and more than 80% reported that it had increased their confidence, job satisfaction and resilience (see Figure 8). Interestingly, one respondent from both ELP1 and ELP2 (but not ELP3) reported a decrease in their job satisfaction as a result of participating in the programme.

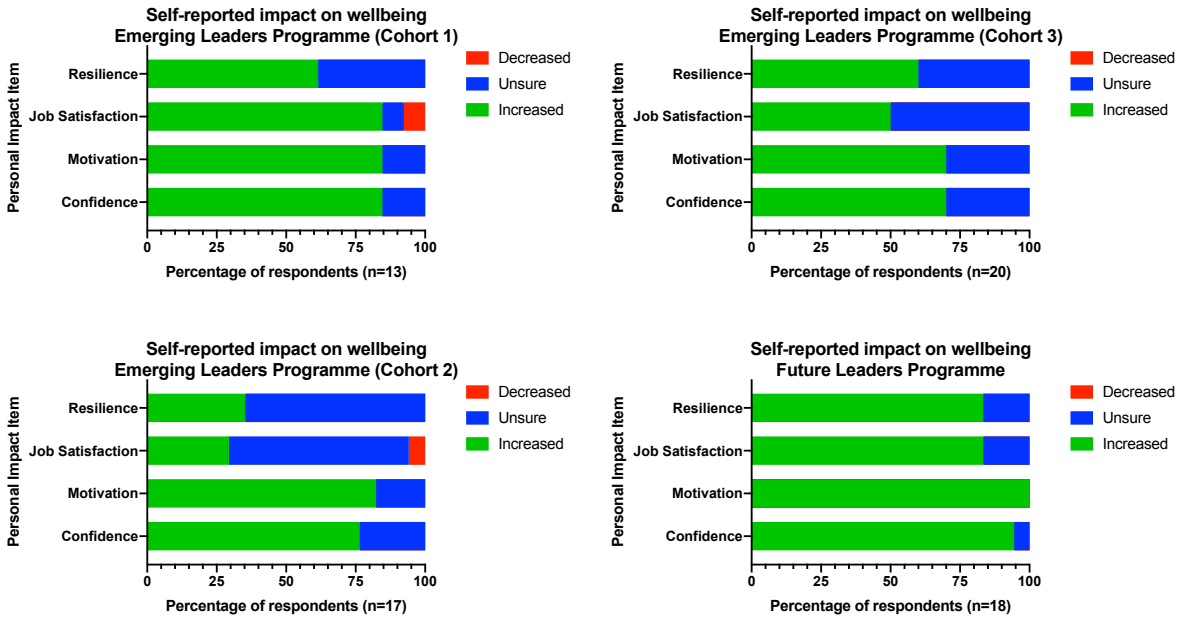


Figure 8: Wellbeing impact of three Emerging Leaders Programme cohorts and the Future Leaders Programme Cohort (2016-2017, bottom right).

As shown in Figure 9, there was no statistically significant change overall in the Brief Resilience Scale in the one cohort it was applied in (ELP3). The mean Brief Resilience Scale score (95% Confidence interval) pre-programme was 22.3 (20.8-23.8) and post-programme was 22.7 (21.2-24.3), with a p-value of 0.6.

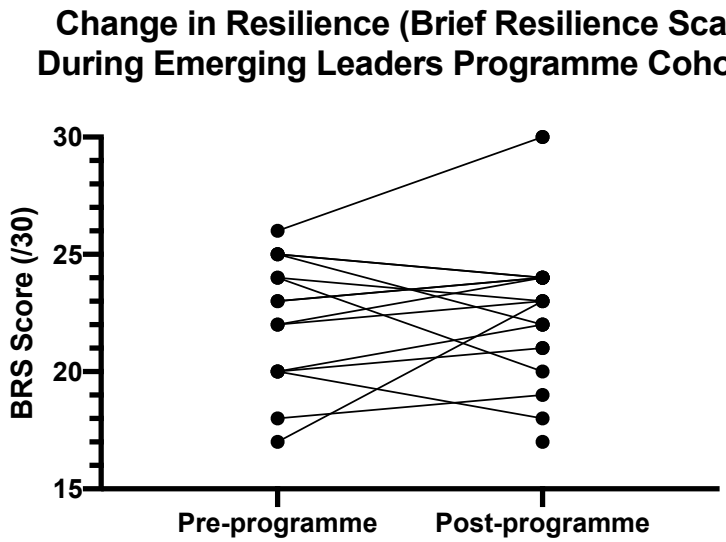


Figure 9: Pre- and post-programme Brief Resilience Scale scores for Emerging Leaders Programme Cohort Three

Free text comments

Participants from all four programmes reported having applied learning from the leadership programme into areas outside the main programme. In clinical settings, participants particularly emphasized their involvement in clinical improvement projects and the way they had built on their learning from the main programme to increase the scale and impact of those extra-programme projects.

- “As I'm going along with being involved in new projects, I'm doing them in a different way, I'm applying my new learning and to a certain extent maybe they end up being bigger, more successful, more impactful because I've started them in a greater position of experience, alongside the FLP project.” (FLP 2016-2017)
- “I guess what I'm try to get at is, there's an awful lot of immediately applied activity which multiplies the impact of the FLP both immediately as well as being the sustainable tools that will continue to influence change over the coming years.” (FLP 2016-2017)
- “Performed a QIP looking at the use of insulin on the Surgical Emergency Unit at the JR. I have performed two cycles and have implemented two changes successfully. This was presented and I was awarded a prize locally.” (ELP1)
- “A lot of improvements instituted in my workplace as initiated by me for patients and staff.” (ELP3)

Participants also noted application into their personal lives, into sports coaching and into national representative roles (FLP 2016-2017). There was less emphasis on these other applications in free text comments, despite most participants having reported such application in other areas in the quantitative questions (see Figure 7).

Confidence

Participants from all four programmes consistently reported the impact of the programme on their confidence. In addition to reporting perceived increases in confidence, participants reported that this change had been noted by others in their workplace, suggesting that their internal confidence had impacted on their behaviour.

- “It's difficult to take credit and say, you know, that's all been just me. But in terms of the confidence of being able to approach, having been a new consultant: which senior people, which key levers, which informatics to grab hold of, and how to demonstrate impact? That's what I've learnt to do, and I'm exercising and still learning about as it applies to other projects in the trust.” (FLP 2016-2017)
- “Gained confidence, and also learnt and built upon existing medical leadership skills.” (ELP1)
- “It enabled me to think about an area outside of my usual comfort zone. It has certainly had a big impact on my confidence and skill set as a project leader.” (ELP2)
- “I am more assertive. My line manager noticed it.” (ELP3)

Motivation

Participants reported that their programme experiences had increased their motivation in their work. This motivation seemed to stem from an increased awareness of the potential for them to have an impact on their environment, as well as the potential for them to develop further as leaders.

- “Gained enthusiasm to be more productive.” (FLP 2016-2017)
- “The ability to see a tangible end product is a motivating factor. Having the means and support to facilitate this from the MDT is encouraging.” (ELP1)

- “Hugely increased motivation to do unconventional things and reflect on what kind of career I want to lead, and what I'm interested in.” (ELP2)
- “I am more motivated to continue focusing on quality improvement projects and developing my leadership skills” (ELP3)

Resilience

Fewer participants directly reported an impact on their resilience in free-text comments. One participant from ELP1 noted that they had experienced challenges in the programme and had developed an increased understanding of how they might approach challenges in their work in the future.

- “Made really good friend in other project partner. The structure of ELP definitely help evaluate our working relationship and be honest with each other.” (ELP1)
- “I think I'm a little more positive about things.” (ELP3)
- “Improved awareness of mental health and wellbeing issues.” (ELP3)
- “I think a lot of what FLP gives is, actually, just, the working together, there is the knowledge that is shared, but actually, what you gain most from it, is the people that you meet, the discussions that you hear, the interactions that you have.” (FLP 2016-2017)

Job satisfaction

Consistent with the quantitative results, most participants reported an increase in their feeling of autonomy at work, and a resultant increase in their job satisfaction. They also reported a change in perspective, with more positivity towards work.

- “Thoroughly enjoyed course, learnt so much, started to apply it in life and have seen good results. Feeling happier at work and making better major

decisions at home, renewed motivation to carry out these kinds of projects at work because now I know there can be some success." (ELP1)

- "It has given me an interesting perspective of the things that I'd love to do, that are sorely missing from my day job. Really eye-opening to change my perspective on my life and career." (ELP2)
- "Great to get to be hands on and do something positive. Have increased feeling of positivity towards medicine." (ELP3)

Negative impacts

Interestingly, a small number of respondents and interviewees reported negative impacts of the programmes. This particularly related to the burden of completing the programme alongside their clinical work.

- First interviewee: "[second interviewee] was telling me the other day, the people who are most motivated probably in terms of programs like this, are the ones who will actually take on all this, and want to do the right thing and probably are more…" Second Interviewee: "Likely to burnout, yeah." (FLP 2016-2017)
- Everybody that I've spoken to is (if we're honest), pretty fed up with still being part of the FLP. Because, you know, it's still a lot of work… quite a lot of people have said you know, I'm ready for it to finish now, so that I can get on with the rest of my busy schedule." (FLP 2016-2017)

Participants did note that the challenges they experienced on the programme were an important part of the developmental experience.

- "I found the lectures great and the project very challenging and often time-consuming. However, on reflection I think the challenge is part of the learning

experience, it's helped me more understand the abilities I have for facing difficulties and overcoming them with the team." (ELP1)

One participant further noted that they felt somewhat disillusioned by their day job after having a change in perspective on the programme.

- "It has given me an interesting perspective of the things that I'd love to do, that are sorely missing from my day job. Really eye-opening to change my perspective on my life and career." (ELP1)

Discussion

The Supplementary Questionnaire was developed to explore the impact of leadership programmes in terms of application of learning outside a programme and in terms of impact on wellbeing. The results show that most participants in the case studies reported application of learning into their clinical work, departmental activities and into their personal lives. They also noted a generally positive effect on their wellbeing, though the responses to this part of the questionnaire varied according to the programme.

Application of learning into clinical and departmental activities and into participants' personal lives is a Kirkpatrick Level Three outcome (subjective change in behaviour) (Hammick et al., 2010; Kirkpatrick, 1979). Experiential learning is generally considered to be a key contributor for effective leadership development (for example, Geerts et al., 2020; Petrie, 2015; Rabin, 2014). The increased confidence and motivation reported across the four case studies could be a causal factor for, or a consequence of, this experiential learning. Increased confidence was a common outcome in the studies included in the systematic review presented in Chapter Five, as it is in the second systematic review presented in Chapter Ten. There are also suggestions in the literature that increased confidence can increase an individual's likelihood of extending themselves into their "zone of proximal development" (Vygotskiĭ & Cole, 1978) and

developing their skills further (Petrie, 2015; Rabin, 2014), so this is an important outcome in itself, particularly given that some participants noted that work colleagues had commented on a change in the participants' behaviour.

The relationship between professional development and increased resilience has been noted elsewhere (for example, Maza, Shechter, Pur Eizenberg, Segev, & Flugelman, 2016; Zwack & Schweitzer, 2013). There is also, however, considerable variation in how individuals define and perceive resilience. The wording of the questionnaire regarding resilience was therefore flawed, as no specific definition was given. Individuals would have needed to apply their own definition of resilience. This is a general problem with single-item construct-level questions, which rely on an individual to interpret the meaning of a question. Such questions have been demonstrated to have a higher risk of bias than behavioural questions (see Chapter Seven for a discussion of single-item construct questions).

The Brief Resilience Scale (BRS) was administered to participants before and after ELP3 as a validated measure of resilience. This score did not significantly change across the programme. It should however be noted that ELP3 started in November 2019 before the Covid-19 pandemic and ended in July 2020, during the Covid-19 pandemic. Decreased resilience and increased burnout was observed in the general population of healthcare professionals over this same period of time (for example, Sultana, Sharma, Hossain, Bhattacharya, & Purohit, 2020). It may therefore be of interest that there was not a significant decrease in resilience during ELP3. Individuals across each of the studies did report positive impacts on support networks and on their perceived quality of life, both of which are considered important for resilience, and could feasibly have helped to maintain participant resilience during the Covid-19 pandemic.

Given the high burden of burnout in healthcare professionals (Embriaco et al., 2007; Maslach & Leiter, 2016; Peckham, 2018; Rowe & Sherlock, 2005) and the impact on

individuals (particularly in terms of health and wellbeing), teams and organisations (for example, turnover, staff sickness, loss of productivity, impact on patient care), it is likely to be worth investing further resources into investigating the possible effects of leadership development programmes on participant job satisfaction and resilience.

Further research applying validated tools such as the Brief Resilience Scale (BRS) (Smith et al., 2008) and the Maslach Burnout Inventory (Maslach & Leiter, 2016) could shed light on this important outcome. Efforts should be made to include control groups in these studies, which could highlight environmental effects that impact both the exposure and control group (such as the Covid-19 pandemic; an organisational crisis; seasonal variation in workload; changes in hospital leadership; participants rotating through jobs).

I was interested to note that a participant in two of the ELP case studies (ELP1 and ELP) reported that their job satisfaction had decreased as a result of the programme. In neither of these cases did the individual explain this in their free-text comments. Several other participants did, however, report increased frustration with their inability to have an organisational impact in their day to day work. Faculty of an unrelated long-standing medical leadership programme based in North England also reported in informal interviews that they have seen participants change roles after dissatisfaction resulting from a new perspective. There are alternative explanations that could also have caused reduction in job satisfaction for these two participants. As mentioned above, there is a well-established high rate of burnout in healthcare workers. It is possible that participants experiencing burnout could have reported reduced job satisfaction, which is a recognised consequence of burnout (Maslach & Leiter, 2016). It is also possible that the increased awareness of other potential roles and future careers which was reported by participants could have resulted in those participants feeling less satisfied in their current role, not because of that role being less satisfying, but because of the increased contrast with more attractive possibilities. Given the investment that many programmes put into

individual leadership development (Beer et al., 2015; West et al., 2015), combined with high burnout levels (Zwack & Schweitzer, 2013), and the relationship between job satisfaction and productivity (for example, Jones, Jones, Latreille, & Sloane, 2008), it is worth considering the possible effects of leadership development programmes on job satisfaction.

As mentioned above and explored in Chapter Seven, single-item construct-level questions risk participants answering the question differently depending on their interpretation. While the Supplementary Questionnaire was reviewed and improved by a programme participant and a faculty member, it would be worth investigating substitution of single-item construct questions in the Supplementary Questionnaire for behavioural questions and conducting cognitive interviews (Willis, 2005) with a pilot group to improve question wording. This could facilitate a more consistent interpretation of questions. Construct-level questions are also more prone to a response-shift bias (see Chapter Seven) which limits their application as a pre- and post- programme. It is therefore appropriate that these questions were asked retrospectively as change-questions.

To test the validity of the Supplementary Questionnaire approach and to test the assumption that participants are able to assess the change in their confidence, motivation, job satisfaction and resilience, it would be necessary to apply a more objective measurement pre-and post- programme, ideally with a control group to assess environmental changes. A control group would help to detect improvement in these measures resulting from regression to the mean, from natural improvement in wellbeing as participants become more experienced in their roles, and from effort-justification bias or an implicitly held theory of learning.

The approach I have taken in this whole thesis is based on Utilisation-Focused Evaluation (Patton, 2008), as outlined in Chapters Two and Three. Utilisation-focused evaluation takes the perspective that evaluation should be focused on the pragmatic

usefulness of the results for decision-making by users. In order to make decisions both about how to evaluate and where to apply efforts to improve the quality of evaluations such as this Supplementary Questionnaire, utilisation-focused evaluation requires that researchers and evaluators collaborate with the primary intended users. In order to assess where efforts to validate and improve this questionnaire would be worthwhile, it would be important to assess the importance of the outcomes measured in terms of decisions to be made by commissioners (scaling or funding the programme), faculty (improving the programme), and participants (whether to go on the programme, whether to recommend the programme to colleagues). Evaluations are rarely without cost, and in order to assess the degree to which their impact on decision making can justify the cost of evaluation, it is vital to consider the perspectives of key evaluation and programme stakeholders. These perspectives will be the focus of Chapter Nine.

Chapter Conclusion

In this chapter I have described the inductive development of a Supplementary Questionnaire which explores both the impact of a programme on a participant in their life outside the programme and the impact on their wellbeing. While there were limitations to this approach, it did uncover a positive general impact on wellbeing and a very high self-reported application of learning into day to day life. Given that leadership development programmes have often been criticised for lack of attention to the transfer of learning into the working environment (for example, Frich, Brewster, Cherlin, & Bradley, 2015; Geerts et al., 2020; Lyons et al., 2020), this is likely to be an important finding.

In order to make decisions as to whether it is worth investing further research into the impact of similar medical leadership development programmes on wellbeing, we need to further explore the importance of these outcomes for key stakeholders. These stakeholders are those who will make decisions of their own regarding funding, improving

and attending leadership development programmes. As outlined in Chapter Two, there are a multitude of explicitly and implicitly held definitions of leadership. It is worth exploring the expectations of stakeholders for leadership development programmes, and the factors which they would consider when making their decisions regarding the programmes, so that further programmes and further evaluation research is able to be optimised.

Chapter Nine: Utilisation-focused evaluation

Chapter Introduction

Research and synthesis of research alike are hindered by lack of agreement as to optimal ways to evaluate programmes. There are a wide range of perspectives and stakeholders in medical leadership development. If we are to improve our collective understanding of optimal medical leadership development, we need to first question our assumptions regarding the purposes of medical leadership development programmes.

As outlined in Chapter Three, I have conducted my research with explicitly practical intentions, and I have adopted a utilisation-focused evaluation approach that considers the goal of evaluation to be to inform decisions. Decisions may be decisions directly relating to the programme: the prospective participant's decision whether to apply for a programme; the faculty member's decisions about which parts of the curriculum to update; the commissioner's decisions as to whether to continue funding a programme. Decisions may also be broader decisions about where to focus further research or where to invest effort in theory development.

Chapters Two and Three highlighted the lack of clarity regarding the definition and meaning of leadership in the context of healthcare. Formal and informal interactions throughout my research outlined to me the variation in interests of groups of stakeholders (commissioners, faculty, participants). These interactions also highlighted that many of these groups were not aware of the implicit interests of other groups.

The aim of this chapter is to begin to explore different ways that different stakeholders decide whether a programme is successful, and to consider the ways in which they could potentially use evaluation to inform their decisions.

This chapter was designed to build on previous chapters with a series of interviews with stakeholders in the main case studies used in this research. Because all of the interviewees worked in healthcare, the advent of the Covid-19 pandemic resulted in most of the planned interviews being cancelled. To compensate for the reduced number of interviews that I was able to conduct, I have revisited other interviews I conducted and have reanalysed their transcriptions looking for relevant themes.

Methods

This chapter focuses on three of the steps that Michael Quinn Patton outlines for effective utilisation-focused evaluation (Patton, 2008). In particular: identify, organise, and engage primary intended users; conduct situation analysis with primary intended users; and identify primary intended uses by establishing the evaluation's priority purposes (Erickson & Noonan, 2018; Patton, 2008, 2013).

Given that the purpose of my research was to develop a rich understanding of the variation in stakeholder uses of evaluation and in their intentions, an inductive approach using interviews was appropriate. My intention was to obtain a rich and varied sample from which to develop initial concepts and to plan future research (Malterud, 2012).

As a result of Covid-19, interviews became less feasible, and I therefore included document analysis as a second key method for this chapter. As such, the key methods used in this chapter were stakeholder identification and analysis, interviews and document analysis, and qualitative synthesis.

Stakeholder identification

NHS Improvement offers a simple framework for identification of stakeholders, as outlined in Table 13.

Group	Description
Commissioners	Those who pay the organisation to do things
Customers	Those who acquire and use the organisation's products
Collaborators	Those with whom the organisation works to develop and deliver products
Contributors	Those from whom the organisation acquires content for products
Channels	Those who provide the organisation with a route to a market or customer
Commentators	Those whose opinions of the organisation are heard by customers and others
Consumers	Those who are served by our customers: that is, patients, families, users
Champions	Those who believe in and will actively promote the project
Competitors	Those working in the same area who offer similar or alternative services

Table 13: The NHS Improvement 9C's framework for stakeholder identification (ACT Academy, 2012).
For the purpose of utilisation-focused evaluation, the organisation can be considered the leadership development programme, and the products could be considered the programme outcomes.

Using a four-quadrant stakeholder power and interest matrix (ACT Academy, 2012; Kammi, 1999; Patton, 2018), I identified key stakeholders or "power actors" (Patton, 2018) for evaluation of these leadership development programmes. This produced an initial list of 130 individuals who were clear stakeholders for the FLP, IMPACT and ELP programmes. After checking this stakeholder map with two experts, a colleague, and a layperson, I concluded that the key stakeholders for this project consisted of those making decisions about the funding for programmes (commissioners), about the delivery of the programmes (faculty, contributors/collaborators), and the participants who apply for programmes (customers).

Sampling

It was not feasible to interview all 130 identified key stakeholders, therefore I adopted a maximum-variation snowball sampling approach. I identified key commissioners for each of the three programmes (FLP, IMPACT, ELP). I then identified a key faculty member from each of the programmes. During the faculty interviews, I asked the interviewee to confidentially suggest participants for whom the programme had appeared

to have most impact, least impact and middling impact. I then approached these participants for interviews.

This sampling process provided three commissioner interviews, three faculty interviews and nine participant interviews, for a total of fifteen planned interviews.

Interviews

Interviews were semi-structured, with a focus on exploring stakeholders' conceptions of success in a leadership development programme, and their intended use of evaluation. I developed a participant information sheet, interview plan and interview guide. A colleague agreed to conduct the interviews for the ELP stakeholders as a research assistant, as I felt my personal involvement in the ELP could create an implicit pressure for them to avoid being critical. With permission from the interviewee, this colleague observed my first interview (IMPACT commissioner) to help ensure consistency of approach across the three programmes. Interviews were planned to last 30-45 minutes. Interviews were recorded and transcribed verbatim using a professional transcription service, and I also wrote notes and reflections during the interviews I completed. Interviewees were invited to read and check transcriptions. I was only able to complete two of the planned fifteen interviews before March 2020, one of which was in person and one by telephone.

Document analysis

After the remaining interviews were cancelled in March 2020, it became clear that I was not going to be able to complete the interviews as planned. Rather than abandoning this key part of my research, I employed document analysis (Bowen, 2009) to collect relevant data by reviewing transcriptions of earlier interviews I had conducted during my research.

Sources for document analysis

I had previously interviewed nine IMPACT programme participants (see Appendix Six – Case study descriptions) to explore the barriers and enablers they had experienced during the programme to applying their learning and achieving tangible results. These interviews also explored the interviewees' perspective on what success of a leadership development programme would mean, and the factors which would support such success. These interviews were all conducted in person in late 2018, lasted for 20-60minutes and were recorded and transcribed verbatim, with notes and reflections written during the interviews and interviewees invited to read and correct transcriptions. While evaluation of programmes was not the only focus of these interviews, analysis of the transcriptions offered relevant data for this chapter. I interviewed five managers, three doctors and a nurse during the IMPACT interviews.

During the development of the Emerging Leaders Programme (ELP), I had interviewed four prospective participants regarding their intended goals for a leadership development programme would be. These interviews were not recorded or transcribed, but I wrote extensive notes and reflections, including quotes from interviewees, and invited interviewees to read and correct the notes. Interviews lasted 20-30 minutes and were conducted in person.

For the Future Leaders Programme (FLP 2016-2017), I had previously interviewed a senior organisational leader who had been faculty on the first cohort and had observed and been involved in commissioning the second cohort. I had written extensive notes and reflections from this interview, as with the ELP interviews. In addition, the same individual had written feedback on the future of the programme.

By combining these previous interviews and documents with the two interviews I was able to conduct, I was able to carry out a preliminary qualitative synthesis.

Qualitative synthesis

I used a process of systematic text condensation as described by Malterud (2012). This involves: first, reading all the material to develop an overall impression and to produce initial themes; second, systematically reviewing the transcription line-by-line to identify meaningful codes (micro-codes); third, condensing and summarising groups of codes to abstract meaning (macro-codes); fourth, interpreting and describing the contents of each group of codes.

Ethics

Ethics exemption was granted by the University of Oxford Clinical Trials Research Governance Committee (CTRG) for each of the case studies as evaluations of service delivery.

Results

Data were collected from a total of 16 individuals across the three case study programmes. These included six managers, three consultant doctors, four medical trainees, one nurse, and two educators.

Macrocodes and microcodes from the qualitative analysis are presented in Table 14 below.

Individuals conceptualised success from leadership development programmes in a range of ways. Many took some time to think through their answers to this line of questioning. Most commented that as a basic standard they would expect to come away from an effective programme feeling “the course had been advantageous to them” (Manager 1, IMPACT). Interviewees routinely went on to explain that an advantageous programme meant that participants had learned either new skills or new ideas “firstly, to develop their leadership and quality improvement skills, and secondly, to broaden their

connectivity and understanding of the connectivity within the trust and understanding of the strategic context within the trust and within the local healthcare economy" (Educator 1, FLP 2016-2017).

Defining Success	Evaluation Considerations
Impressions	Complexity
Positive feelings	*No clear causation*
Recommendations to colleagues	Flexible goals
Self-fulfilment	*Participant-set goals*
Confidence	*Emergent goals*
New Skills	Timeframe
Soft skills	*Change takes time*
Financial skills	*Short-term outcomes*
Quality improvement skills	*Long-term transformations*
New Ideas	
Understanding of Leadership	
Personal Mastery	
Self-awareness	
Changing behaviour	
Applying learning into work	
New leadership roles	
Improved teamwork	
New connections	
Results - individual	
CV building	
Results - team	
Impact on others	
Project Outcomes	
Results - organisational	
Cultural Change	
Return on Investment	
Improved leadership capacity	

Table 14: Results of the systematic text condensation. Macrocodes are listed with microcodes in italics.

Many individuals suggested that they would expect a successful programme to result in participants applying their new skills and ideas into their work as a change in behaviour, to "[make] that transition/change to this leadership role which is their daily work" (Trainee 1, ELP).

There was also a strong recurring focus on how changes in how participants behaved might have an impact on their teams and for their patients. "In the short term the hospital obviously benefits from a project with successful outcomes both in a patient sense as

well as a fiscal sense. Individuals also benefit from seeing the effect of their work" (Commissioner, IMPACT).

While a range of individuals did mention a financial return on investment as a marker of success, this was a source of divergence amongst stakeholders. Some individuals rather cautioned that financial return on investment should not be seen as the most important outcome from a programme: "I think the great risk is that that becomes the all-around measure of success, when that is misconstruing the purpose" (Educator 1, FLP 2016-2017).

Interviewees suggested that it is important that there is flexibility in how a programme's success is measured, and that it is important for participants to be able to select their own goals and measure their own progress. "I think that everybody is on the same journey so everybody you know, success is determined by the outcome. You know your goals you're trying to achieve" (Manager 2, IMPACT).

There was considerable attention paid to the complexity of leadership development, and the difficulty in measuring outcomes and in attributing outcomes to a programme. One participant listed important outcomes, and simply stated after each that they weren't measurable: "Improved teamworking, not measurable. Improved self-fulfilment, not measurable…" (Consultant 1, IMPACT). The difficulty in measuring outcomes and in attributing causation was seen to be a barrier to decision-making by some of the individuals, because of the way people look for objective results. "I think it's important to acknowledge that there's not going to be something that looks like scientific proof or anything close to it" (Educator 1, FLP 2016-2017).

While others suggested that an overall impression of the organisational impact of a programme would arise, they emphasised that this would take years to be seen and would require patience from commissioners and faculty who would want to see short-term outcomes. "The impact thing will be under pressure, in terms of how can it change

and show results in the year. But I don't believe that it can… We want more people from other places to talk about leadership in general. And then for us to just sit on that" (Manager 1, IMPACT).

While the interviews and document analysis brought in a relatively small sample, there was divergence of opinions regarding several factors. Prospective participants all mentioned CV-building as an important outcome, whereas none of the commissioners did. Faculty and commissioners tended to cite networking and community building as important outcomes, and while some participants mentioned these outcomes, most focused on learning knowledge and skills. Further research would be needed to explore how different stakeholders vary in the outcomes that they consider to be important to evaluate.

Discussion

This exploratory analysis suggested that there are a range of conceptions of success prevalent across even this small sample. Emphasis was placed by participants, faculty and commissioners alike on new skills and ideas and on tangible results, but there was little clarity as to how these skills, ideas and results could be feasibly and credibly measured and communicated. The need for patience was highlighted, with a range of individuals commenting that it would take years for the real effects of an initiative to be seen.

The results of the systematic text condensation was more clearly aligned with Avolio's framework of affective, behavioural, cognitive and organisational outcomes (Avolio, Reichard, et al., 2009) than with Kirkpatrick's Framework (Kirkpatrick, 1967). Many participants commented on new ways of thinking being important for an effective programme, a set of outcomes which do not easily fit into Kirkpatrick's framework.

Some of the codes, such as CV building, did not clearly fit within either framework. This emphasises that current frameworks remain inadequate to encapsulate important outcomes, and aligns with the established position in utilisation focused evaluation that it is important to explore specific stakeholder expectations and intentions and to align evaluation to these (Hannum, 2004; Hannum et al., 2007; Patton, 2008; Vassar et al., 2010).

It was interesting how much effort interviewees seemed to need to put into thinking through what a successful outcome might mean to them. While the educators seemed to have relatively quick and articulate responses, participants often paused or took some time before they were able to articulate a clear statement. Given the variety of explicit and implicit definitions of leadership (see Chapter Two) this should perhaps not be surprising. Many of the participants were at the end of a leadership development programme, having attended several leadership workshop sessions, but were unable to immediately articulate what success in a leadership development programme would look like. This could suggest that clearer conceptions of leadership were not important to these individuals; that they were not able to develop clearer conceptions during that specific programme; or that the fact that a specific programme focused on leadership was not, in fact, the reason why they attended the programme. Alternatively, the fact that participants could not easily articulate what success in a leadership development programme would look like, could suggest that the programmes failed to provide participants with a clear understanding of what success could be.

The work described in this chapter was undertaken to clarify how participants and funders make decisions about medical leadership development programmes. The apparent lack of individual clarity suggests that this is an important area for further research.

Patton (2018) provides some initial suggestions as to how various stakeholders broadly intend to use evaluations. In lieu of richer interview data, these may offer a useful starting point. He suggests that those who fund programmes and write policies, will want to “render judgements of overall merit, worth, and significance”. Programme staff will be most interested in “formative evaluations aimed at program improvement”. Scholars and academics will be interested in “knowledge-generating and theory-driven evaluations” (Patton, 2018).

Leadership development programmes usually require substantial financial and time investments (Avolio et al., 2010). Evaluations also can require substantial investments of time and money (Hannum et al., 2007). If evaluations are to be effective in informing decisions regarding investments of time, money and effort into leadership development programmes by organisations, commissioners, faculty and participants, we need to make further efforts to ensure that they are targeted to specific, reliable, credible and valid measures (Hannum et al., 2007; Patton, 2008).

Chapter Conclusion

Overall, the results of this chapter suggest that early and thorough identification of and engagement with stakeholders is vital for effective utilisation-focused evaluation. There may be value in conducting more extensive interviews and surveys regarding medical leadership development, and in characterising the most important evaluation outcomes for each group of decision-makers across a range of medical leadership development programmes. By collating this information and mapping to evaluation strategies, we would be able to provide future evaluators and researchers with a useful starting point for planning and undertaking their own evaluations.

Part III of this thesis has focused on evaluation of medical leadership development programmes for qualified doctors. Because this constituted a shift from Part II, which

presented a gap in medical leadership development for medical schools, I conducted a second systematic review, which I will present in Part IV of this thesis.

PART IV:

Systematic Review of Medical Leadership Development Programmes for Qualified Doctors

Chapter Ten: Evidence-Based Medical Leadership Development: A Systematic Review

Note on publication of Chapter Ten

The research I have presented in this chapter was published in BMJ Leader as Lyons, O., George, R., Galante, J., Mafi, A., Fordwoh, T., Frich, J., & Geerts, J. (2020). Evidence-based medical leadership development: a systematic review. *BMJ Leader*. https://doi.org/10.1136/leader-2020-000360. In Chapter Ten I have adapted this published paper to make it compliant with departmental regulations for including published work in a doctoral thesis, and to make it clear how this work relates to other chapters in this thesis.

Contribution statement

As the lead author for this paper, I conducted all aspects of this research, including study design, search strategy, screening and selection of included papers, coding, quality appraisal and data analysis, drafting, editing and approval of the final manuscript.

For the purpose of publication, and to comply with expected standards for systematic reviews, other members of this project team (Robynne George, João Galante, Alexander Mafi and Thomas Fordwoh) duplicated parts of the research in parallel, as outlined in the text below.

All co-authors approved the final manuscript, and all have approved the inclusion of this research in my thesis.

Chapter Introduction

In Part II (Chapters Four and Five), I focused on leadership development for medical students and I established a gap in medical leadership development, both in Australia and New Zealand and internationally.

In Part III (Chapters Six, Seven, Eight and Nine) I moved from a focus on medical students into practical development of evaluation methods for medical leadership development programmes for qualified doctors. I explored the development of the Revised Primary Colours Questionnaire. I analysed the modification of the Revised Primary Colours Questionnaires to investigate response-shift bias with retrospective then-tests. I described the development of the Supplementary Questionnaire. Finally, I explored the divergence of expectations that various decision-makers have in terms of evaluating medical leadership development programmes.

In Part IV (Chapter Ten), I have presented the results of a second systematic review. There are two main reasons for conducting a second review.

First, while Part II of my thesis focused on gaps in leadership development in medical schools, the case studies I was able to study in Part III focused on qualified health professionals, and therefore I felt that I needed to conduct a new systematic review covering later career stages.

Second, Part III of my thesis reiterated the need for validated evaluation methods and effective questionnaires and raised questions about the reliability of existing research using existing flawed evaluation methods. While the systematic review in Chapter Four used a broad critical appraisal tool, I wanted to more extensively appraise the quality of medical leadership development programme research and to highlight common features of more reliable evaluations.

A previous systematic review of 45 studies evaluating leadership development interventions for doctors found that effective interventions were characterised by the use of multiple learning methods, including seminars and group work, alongside action learning projects in multidisciplinary teams (Frich et al., 2015). These findings were echoed in a recent study by Geerts et al. (2020), who emphasised that plans need to be in place for transferring learning from the intervention into the working environment

Evidence backing medical leadership development activities has been variable in quality (Clark & Armit, 2008; Frich et al., 2015; Geerts et al., 2020; Leslie et al., 2005; Rosenman et al., 2014; Sadowski et al., 2018; Steinert et al., 2012; Straus et al., 2013; West et al., 2015). There has been a particular lack of research and evaluation that goes beyond individual learner feedback and subjective outcomes (Frich et al., 2015; Geerts et al., 2020; Husebø & Akerjordet, 2016; Rosenman et al., 2014).

We undertook this second systematic review to synthesise recent empirical evidence regarding medical leadership development programme factors associated with outcomes at the clinical and organisational level. We specifically investigated links between aspects of programme design, delivery and evaluation and improved outcomes.

Given the variable quality of studies highlighted in previous reviews (Frich et al., 2015; Geerts et al., 2020; Rosenman et al., 2014), we applied two validated critical appraisal instruments (Aromataris & Munn, 2019; Reed et al., 2007) to isolate higher-reliability findings. This review is the first to apply both instruments in order to identify and synthesise the highest quality empirical evidence in medical leadership development.

Methods

The design of this review was guided by the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) (Liberati et al., 2009) and the Best Evidence in Medical Education (BEME) guide for systematic reviews (Hammick et al., 2010). Our

methods were based on the review conducted by Frich et al. (2015), with methodological changes drawn from other reviews (Geerts et al., 2020; Husebø & Akerjordet, 2016; Rosenman et al., 2014; Sadowski et al., 2018; Steinert et al., 2012; Straus et al., 2013). Following the BEME recommendations for systematic reviews (Hammick et al., 2010), we hand-searched references and citations of known reviews extensively to supplement our database search. In line with recommendations from Geerts et al. (2020) and Rosenman et al. (2014) we assessed study quality using the Medical Education Research Study Quality Indicator (MERSQI), which is designed to measure the methodological quality of quantitative medical education research studies (Reed et al., 2007). We added the Joanna Briggs Institute (JBI) Critical Appraisal Checklist (Aromataris & Munn, 2019), which is designed for meta-aggregation of qualitative research and is well-established in healthcare research (Lockwood et al., 2015).

Search strategy

We began this review by re-examining the data set identified in the review of leadership development for physicians by Frich et al. (2015). With assistance from a specialist librarian at the University of Oxford, we then based our search strategy on Frich et al.'s review (2015). Using the Ovid MEDLINE database, we conducted a search using both free text and Medical Subject Headings (MeSH). The full search terms are listed in the supplementary material. This search identified 501 unique publications. We then conducted an extensive hand-search of references and of citations in known medical leadership development reviews using Web of Science and Google Scholar. This identified an additional 107 studies for possible inclusion, for a total of 608 records for screening (Figure 10).

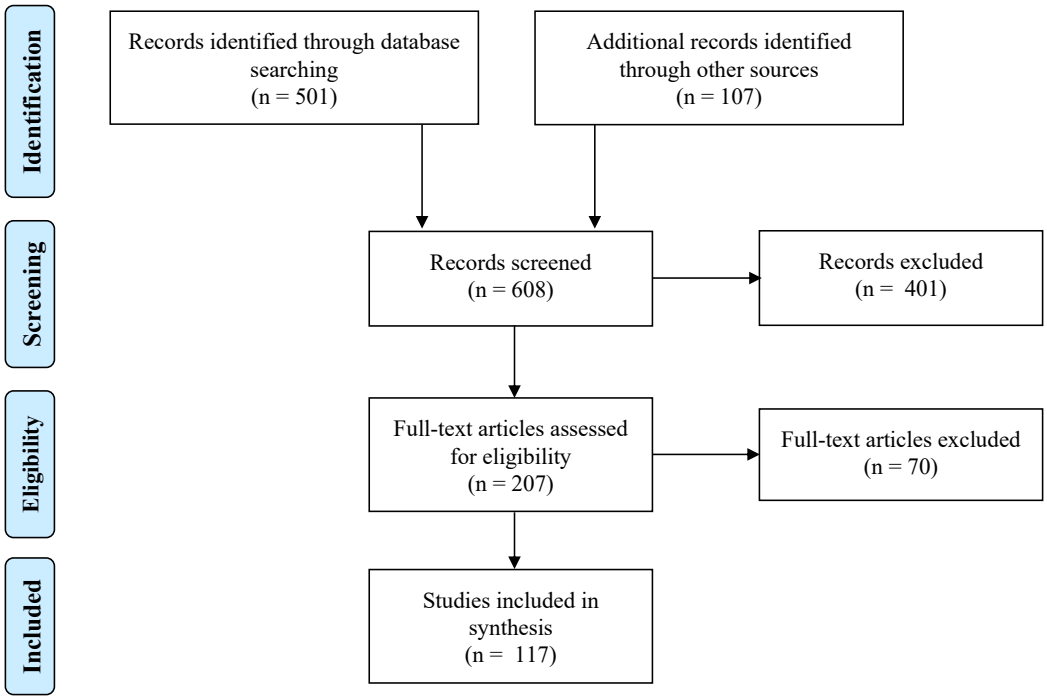


Figure 10: PRISMA Diagram

Inclusion criteria

We included any peer-reviewed study published in English between January 2000 and January 2020 which:

1. Describes a leadership development intervention (programme, workshop, course, etc.);
2. Includes physicians as learners (defined here as any practicing doctor post-qualification);
3. Evaluates the leadership development intervention.

Qualitative, quantitative, and mixed evaluations were included. We excluded studies where leadership development was a minor focus or where the proportion of physicians was lower than ten percent of intervention participants.

Screening process

Two members of the review team (OL and TF) reviewed study titles and abstracts for eligibility. Articles that were approved by either reviewer progressed to full-text review. Two members of the review team independently reviewed for inclusion the full text of all 207 articles that passed the title and abstract screen (OL reviewed all, TF and RG reviewed half each). Where there was disagreement about inclusion, all three reviewers (OL, TF, RG) reached consensus by discussion, with the third reviewer (TF or RG) arbitrating where required.

Data abstraction

After screening and reviewing for eligibility, 117 unique studies were included for abstraction and analysis. Data were extracted and coded for educational setting, methods, content, and for evaluation methods and outcomes. Outcome data were categorised according to an adapted version of Kirkpatrick's Framework for Evaluation of Training Programmes (see Table 15) (Hammick et al., 2010; Kirkpatrick, 1979). One reviewer extracted and coded all 117 included studies (OL). The second reviewers (RG/JRG/AM/TF) each extracted and coded at least five studies in full to ensure consistency between reviewers. Data extraction and coding for all 117 studies was then cross-checked by the second reviewers. Any differences were resolved by consensus, with a third reviewer arbitrating where required. Where possible, statistical tests performed in studies were replicated and checked for accuracy.

Kirkpatrick Level	Description
Level 1 Reaction	Participants' satisfaction with the learning experience, its organisation, presentation, content, teaching methods, and quality of instruction
Level 2A Change in Attitudes	Changes in the attitudes or perceptions among participant groups towards leadership, management, and/or administration
Level 2B Change in Knowledge or Skills	For knowledge, this relates to the acquisition of concepts, procedures, and principles; for skills, this relates to the acquisition of thinking/problem-solving, psychomotor, and social skills
Level 3A Behavioural change (self-reported)	Documents the transfer of learning to the workplace and changes to professional practice, as noted by participants themselves
Level 3B Behavioural change (observed)	Documents the transfer of learning to the workplace and changes to professional practice, as noted by a third party or by promotions
Level 4a Results (self-reported)	Organisational results perceived by respondents and group effectiveness perceived by subordinates
Level 4b Results (observed)	Tangible organisational results, such as reduced costs, improved quality and safety, impact of projects

Table 15: Kirkpatrick's Framework For Evaluation of Training Programmes, with adaptations from Frich et al. (2015).
This adaptation of Kirkpatrick's Framework varies slightly from the versions in Chapter Four and Chapter Three.

Study quality appraisal

Previous reviews have shown marked variation in the quality of studies of medical leadership development (Geerts et al., 2020; Husebø & Akerjordet, 2016; Rosenman et al., 2014; Sadowski et al., 2018; Steinert et al., 2012; Straus et al., 2013). To isolate the most reliable evidence linking medical leadership programmes to improved outcomes, two researchers (OL and RG/JRG/AM/TF) independently critically appraised each included study using the MERSQI and JBI Instruments (Lockwood et al., 2015; Reed et al., 2007). Differences in MERSQI and JBI quality score were resolved by consensus, and a third researcher arbitrated where needed.

The MERSQI was applied to all articles. The MERSQI is a validated appraisal tool consisting of ten items in six domains which relate to design, sampling, type of data collected, validity of evaluation methods, analysis and outcomes (Reed et al., 2007). Each domain is scored to a maximum of three, for a total score of 5-18. In line with Geerts

et al. (2020), studies with scores of twelve or higher were categorised as higher reliability studies (see data analysis section below).

The Joanna Briggs Institute (JBI) Checklist for Qualitative Studies was also applied where an article used mixed methods (k=53) or qualitative methods (k=10). Fundamental differences in study design, sampling, evaluation instruments, and analysis preclude summative comparison of mixed-methods or qualitative studies to quantitative studies using the MERSQI (Côté & Turgeon, 2005; Hannes et al., 2010; Lockwood et al., 2015; Reed et al., 2007). The JBI checklist is considered the most appropriate qualitative critical appraisal tool for use in pragmatic meta-aggregation of qualitative research (Hannes et al., 2010). It includes ten items which regard the study's research questions, methods, analysis and reporting, for a total score of 0-10. Following recommendations from the JBI Reviewers' Manual (Aromataris & Munn, 2019), a cut-off score for higher reliability studies was predetermined at 6/10. This score was chosen as studies obtaining six or more points included most key elements of high-quality design.

Data analysis

MERSQI and JBI scores were used to establish which studies presented more reliable evidence of outcomes. Summary statistics were calculated for all 117 studies. In line with Geerts et al. (2020), studies with a final MERSQI score of 12/18 or higher were also analysed separately to isolate the most reliable evidence, as were qualitative and mixed-methods studies which achieved the pre-determined JBI score of 6/10 or higher. As there was substantial methodological heterogeneity, study characteristics and outcomes were synthesised using a mixed-methods approach (Munn et al., 2014; Pearson et al., 2015). All study quality appraisal scores are presented in the supplementary material in Appendix Two, and full data extraction tables are available on request.

Results

Study reliability (MERSQI and JBI)

28 of 117 studies (25%) were categorised as higher reliability. Two studies were categorised as higher reliability by both the MERSQI and the JBI tool (Cooper, 2001; Orme & Campbell, 2019), 14 studies (12%) by the MERSQI only and 12 studies (10%) by the JBI tool only. The median critical appraisal score according to the MERSQI was 8.5 (range 5-16 from possible range of 5-18) and the median critical appraisal score according to the JBI was 3 (range 0-9 from possible range of 0-10). Table 17 (see Appendix Two – Supplementary data for Chapter Ten) includes the MERSQI and JBI scores for all included studies.

Study design showed considerable room for improvement, as shown in Table 18 and Table 19 (see Appendix Two – Supplementary data for Chapter Ten). Nearly half of the included studies (46%) relied on post-programme evaluations only, and 92% did not include a control group. Of the nine studies that did include control groups, most had substantial methodological flaws in their selection of control groups. One common method for control group recruitment was to use unsuccessful programme applicants (Berkenbosch et al., 2014; Day et al., 2011; Fassiotto et al., 2018). In terms of evaluation design, the median evaluation instrument score was zero (range 0-3). Most studies (59%), did not fulfil any of the MERSQI requirements for evaluation instruments, including reporting questionnaire design, wording and content. Objective outcome measures were used in only a minority of studies, with 60% relying solely on self-reported measures.

Data analysis and reporting likewise showed considerable limitations. Only one in five studies (20%) met criteria for comprehensive analysis and reporting of data. Few studies analysed their data beyond descriptive statistics to consider the generalisability and implications (13%). In many cases, studies omitted basic statistical significance tests.

Many studies did not contain key reporting elements for qualitative research as outlined in the JBI tool (see Table 19 in Appendix Two – Supplementary data for Chapter Ten). There was clear congruity between research methodologies chosen and the research objectives and methods employed in 60% of studies. A minority of studies adequately reported their analysis (28%) and interpretation of data (25%), the potential for the researcher to have influenced data collection and interpretation (23%), and the researcher's cultural or theoretical orientation (15%). Participant voices were clearly represented through quotes in only 16/53 (30%) of mixed methods studies and 5/10 (50%) of qualitative studies. There was a statement of ethical approval or ethics exemption in only 26 of 63 studies (40%) which used qualitative methods. No study included a statement of philosophical perspective (normally expected for qualitative research) (Aromataris & Munn, 2019).

Programme design

There was considerable heterogeneity in leadership development intervention design. It was often unclear whether established good practice for development of medical education interventions (Geerts et al., 2020; Thomas et al., 2016) was followed, as shown in Figure 11. Only 52 studies (44%) reported having conducted a needs assessment before their intervention, and only 20 studies (17%) explicitly reported using an established capability or competency framework to inform leadership programme goals and objectives. There was, however, a plan for training transfer reported or built into 68 of 117 interventions (59%).

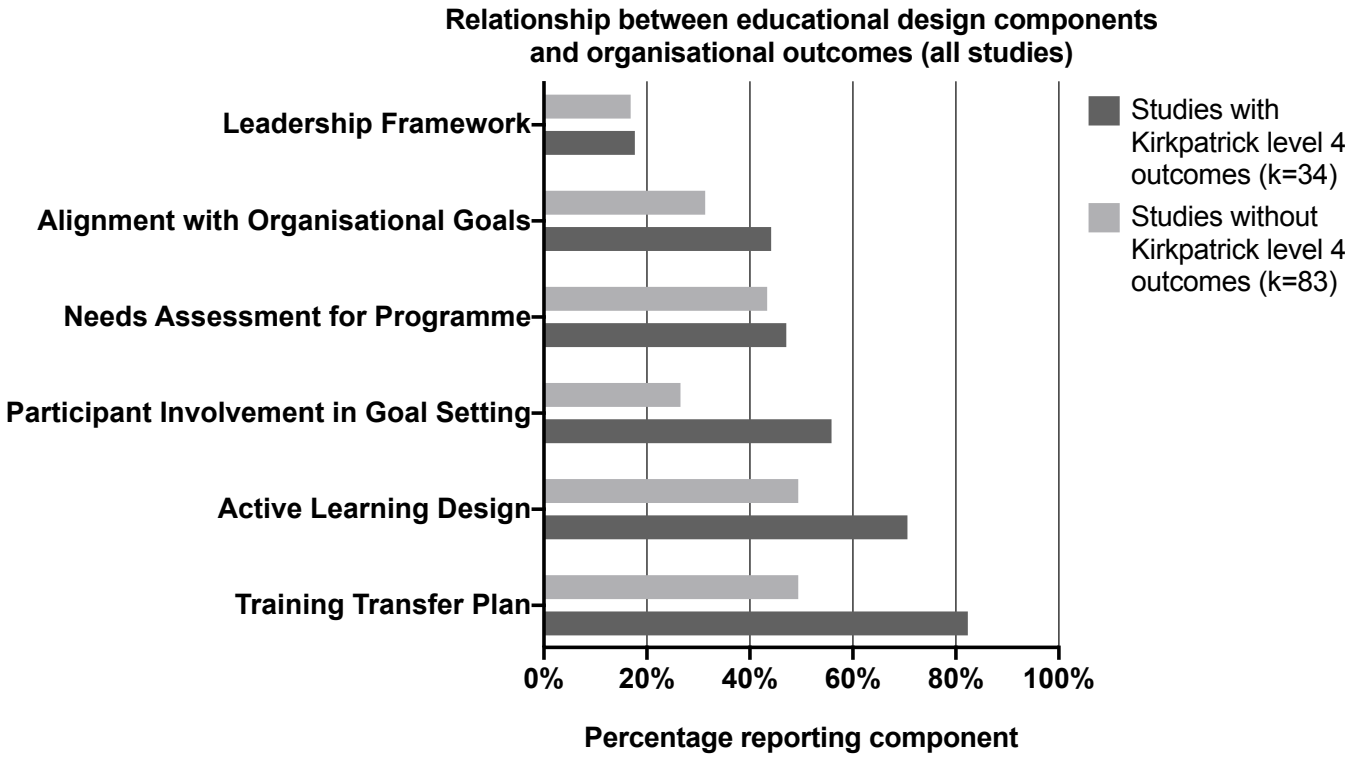


Figure 11: Educational Design Components
Studies which reported Kirkpatrick Level 4 Outcomes (k=34) compared with studies that did not report Kirkpatrick Level 4 outcomes (k=83).

The majority of interventions were carried out in a single hospital department (27%), single hospital (22%), or a single university (12%). Just under a quarter (23%) of interventions were conducted in multiple healthcare centres, and a further 15% of studies were conducted within a specialty training programme outside healthcare centres.

Most of the studies took place in the United States (67%) or the United Kingdom (16%). The remainder of studies were in other European countries (7%), Canada (4%), or Australia (3%), with a single study each from Africa (Nakanjako et al., 2015), India (Gulati et al., 2019), Israel (Maza et al., 2016), and Qatar (Al-Mutawa et al., 2016).

Programmes ranged in length from two hours to four years. The median intervention length was six months, and the most common length was one year (19%). Only 18 interventions (15%) lasted longer than one year. Five interventions (4%) were shorter than one day.

Programme faculty

Programmes were predominately delivered by either in-house faculty (36%) or a mix of in-house and external faculty (32%). Programmes delivered by a mix of faculty were most likely to show organisational outcomes, as shown in Figure 12 (p=0.049 using chi-square test). The professional backgrounds, qualifications and experience of faculty were generally not reported.

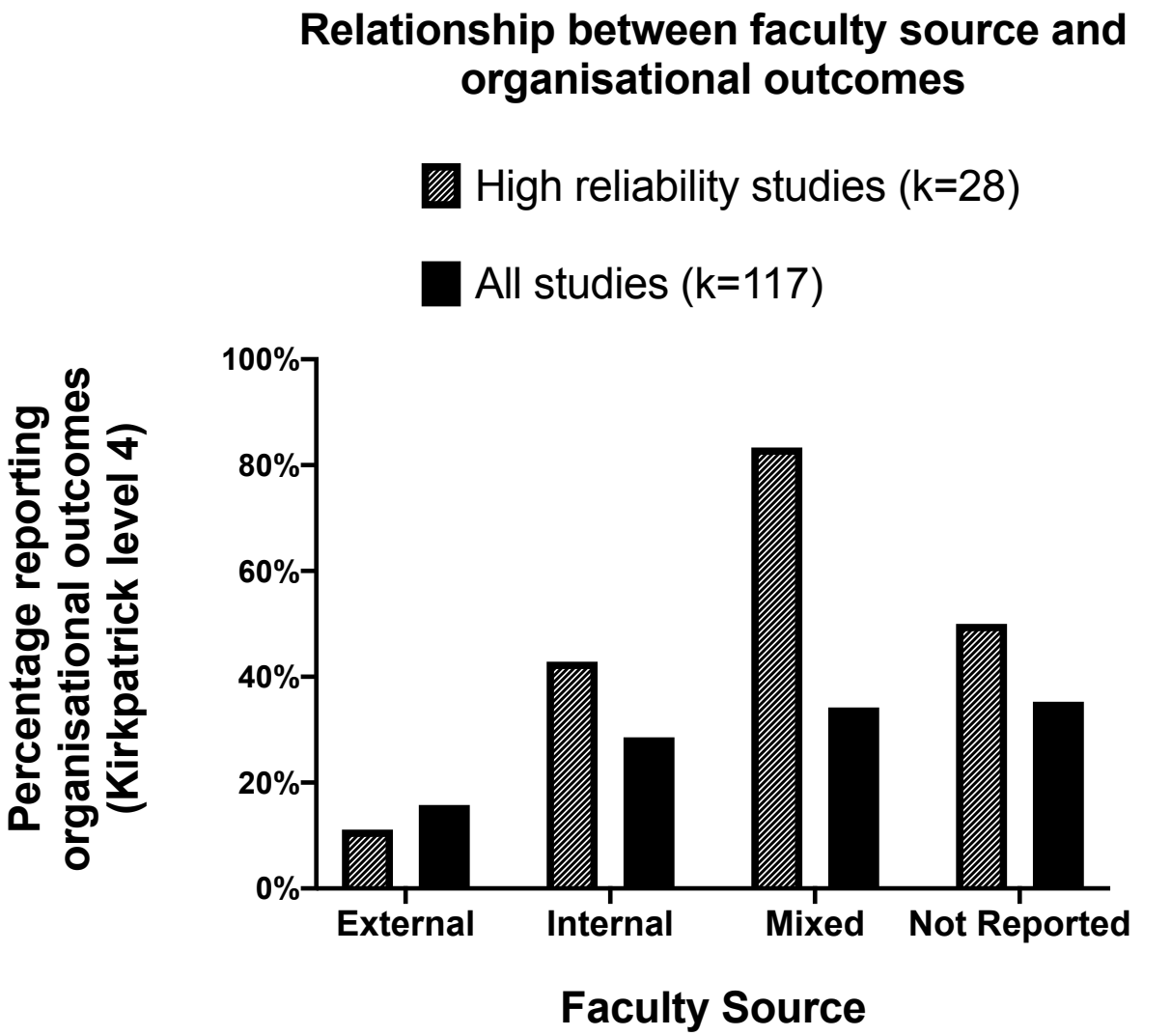


Figure 12: Relationship between faculty source and programme outcomes.
Higher reliability studies were those with Medical Education Research Study Quality Indicator Score of at least 12/18 or Joanna Briggs Institute score of at least 6/10. NR= Not Reported

Participants

The majority of programmes included doctors only (76%). Doctors ranged from residents (60%) to full specialists (30%) and academic medical faculty (19%). Only nine

studies of 117 involved doctors from more than one of these experience levels. Behavioural outcomes were reported in a similar percentage of higher-reliability studies for each of these experience levels (85-92%). In contrast, organisational results were more commonly reported in programmes with academic medical faculty (50%) or full specialists (44%) than in programmes with only residents (20%).

The 26 studies (24%) reporting multidisciplinary programmes included a combination of nurses (12%), managers (15%), and allied health professionals (9%). Most studies did not report the gender of participants (74%) or the age of participants (87%).

In terms of participant selection criteria, the majority of interventions included participants who volunteered (27%), were nominated (19%), or who applied to the programme (16%). In some cases, the application process was highly competitive. Interventions were mandatory in one fifth of studies (20%). A considerable proportion of all studies (23%) did not report the selection process for their learners, including one quarter (25%) of the studies categorised as higher reliability by MERSQI criteria.

Educational methods

A wide range of educational methods were employed in various combinations across the reviewed studies, as shown in Figure 13. Most interventions included lectures (68%) and small group work (61%). Project work was included in the majority of studies with organisational outcomes (68%), but only in a minority of studies which did not report organisational outcomes (33%). Individual or team mentoring was also more prevalent where organisational outcomes were reported (47% vs 23%).

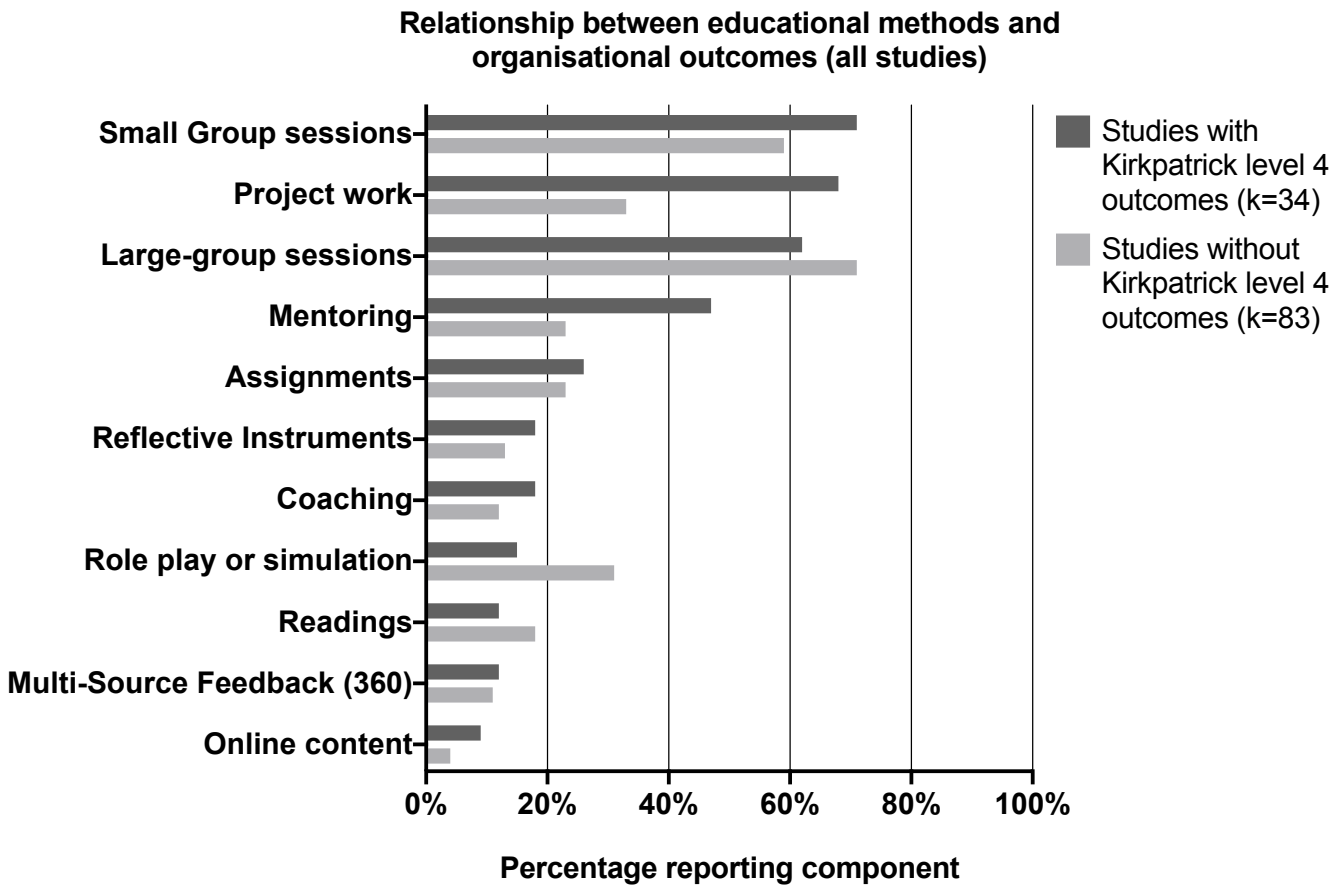


Figure 13: Educational Methods
Studies which reported Kirkpatrick Level 4 organisational outcomes (k=34) compared with studies that did not report Kirkpatrick Level 4 organisational outcomes (k=83).

Educational content

Educational content varied considerably among interventions. The most consistent content area was leadership theory (reported in 65% of interventions). The other common content areas were performance management (44%), self-management (41%), change management (39%), communication (36%), teamwork (33%), quality improvement (30%), healthcare policy (27%), healthcare finance (26%), and leadership behaviours (20%). There were no notable educational content differences in higher reliability studies or in studies which reported organisational outcomes (Kirkpatrick Level Four).

Evaluation methods

A wide range of evaluation methods were employed across the included studies. Nearly half utilised quantitative methods only for their evaluation (46%). Of the

remainder, most studies used mixed methods (45%), with ten studies (9%) using purely qualitative methods. These proportions were similar in the higher reliability studies (41% quantitative, 48% mixed methods, 10% qualitative).

Four out of every five studies (82%) used questionnaires in their evaluation. Almost all of these studies employed Likert scale items (92%); one third included open questions (34%). Only nine studies (8%) used validated questionnaires. The proportion of higher reliability studies using validated questionnaires was greater, though still a minority at 20% (MERSQI) and 18% (JBI). An additional six studies (6%) had conducted an expert review of their questionnaire for content validity only.

More than two-thirds of the included studies relied solely on self-ratings (69%). A minority of studies included ratings from subordinates (3%), peers (7%), superiors (12%) or experts (20%). The proportion of higher-reliability studies which relied on self-ratings was lower (39%), with increased use of ratings from peers (14%), superiors (25%) or experts (39%).

The majority of studies (72%) included the collection of outcome data regarding behavioural changes (Kirkpatrick Level Three, 57%) or organisational outcomes (Kirkpatrick Level Four, 24%). Only three studies relied solely on Kirkpatrick Level One outcomes (reaction) (Bhatia et al., 2015; Johnson & Stern, 2014; Ringdahl et al., 2014).

Nearly half of the studies utilised single group post-programme only designs (46%), with most of the other half using single group pre- and post-programme designs (46%). Most studies included a post-programme evaluation completed immediately at the end of the programme (90%). Only 18 studies (15%) included a longer-term evaluation. In higher-reliability studies, longer-term evaluations were associated with increased reporting of organisational results (56%) when compared to immediately-post designs (31%). All 16 higher reliability studies as assessed by the MERSQI used pre- and post-designs. Six of these included a non-randomised control group (38%) and one study

included a randomised control group (6%). This was the only randomised control group used in any of the 117 studies.

Behavioural and organisational outcomes in higher reliability studies

A full summary of outcomes from all 117 studies is provided in Table 17 (see Appendix Two – Supplementary data for Chapter Ten).

There was a range of behavioural (Kirkpatrick Level Three) and organisational (Kirkpatrick Level Four) outcomes demonstrated in higher reliability studies.

Behavioural changes were objectively demonstrated in higher reliability studies through observed changes in behaviour (Boyle & Kochinda, 2004; Cole et al., 2017; Cooper, 2001; Gilfoyle et al., 2007; Orme & Campbell, 2019; Ruston & Tavabie, 2010; Ten Have et al., 2013), promotions (Dannels et al., 2008; von Vultée & Arnetz, 2004), increased responsibilities or titles (Agius et al., 2015; Fassiotto et al., 2018; Haftel et al., 2018; Mckimm et al., 2019; Tsoh et al., 2019) and project completion (Hopkins et al., 2018; Levine et al., 2008; Throgmorton et al., 2016). Subjective changes in behaviour included improved communication (Boyle & Kochinda, 2004), influence (Hopkins et al., 2018), delegation (Orme & Campbell, 2019), collaboration (Pradarelli et al., 2016), involvement in service improvement (Mckimm et al., 2019), and application of skills learned or improved leadership in general (Bergman et al., 2008; Boyle & Kochinda, 2004; Cohen et al., 2017; Monkhouse et al., 2018; Ruston & Tavabie, 2010; Wurster et al., 2007). These changes were indicated through interviews, free text questionnaire responses and behavioural self-assessments.

Organisational outcomes in higher reliability studies (Kirkpatrick Level Four) were defined prospectively and in most cases were objectively demonstrated through leadership project impact evaluations. Projects achieved a range of outcomes, including reduced waiting times (Hopkins et al., 2018), improved patient care (Agius et al., 2015;

Hopkins et al., 2018) and cost savings (Agius et al., 2015; Hopkins et al., 2018; Mckimm et al., 2019; Orme & Campbell, 2019). By assessing the financial impact of projects completed during the intervention and relating this to programme costs, one higher reliability study reported a 364% financial return-on-investment (ROI) (Orme & Campbell, 2019). Other objective outcomes included reduced organisational turnover of participants (Fassiotto et al., 2018), improved departmental working climate (Boyle & Kochinda, 2004), reduced sick leave (von Vultée & Arnetz, 2004) and increased promotion of women (Dannels et al., 2008). Organisational outcomes were subjectively indicated through reports of increased staff retention (Monkhouse et al., 2018) and improvement in organisational effectiveness (Orme & Campbell, 2019). One study reported that “intangible benefits” resulted in a 106% financial ROI (Throgmorton et al., 2016).

Organisational outcomes in higher reliability studies were reported more frequently from programmes delivered by a mix of internal and external faculty than by only external faculty (83% vs. 11%), as shown in Figure 12 above. Organisational outcomes were also more frequently reported from interventions conducted in a whole hospital (57%) or multiple hospitals (40%), compared to interventions conducted in a single specialty (conference or outside-hospital training programme) (33%), single university (25%), or in a single department (0%). There were no notable differences in outcomes related to specific educational content.

Higher reliability studies that reported organisational outcomes were more likely have included project work (70% vs 44%), mentoring (50% vs 22%), coaching (22% vs 11%), and reflective instruments such as personality type assessments (22% vs 6%) than higher reliability studies that did not report organisational outcomes. Organisational outcomes were reported less frequently in higher reliability studies that included simulation or role play (10% vs 33%).

Discussion

The main aim of this review was to synthesise recent empirical evidence regarding evaluation of medical leadership development programmes for qualified doctors.

We found a substantial increase in the number of studies which evaluate medical leadership development interventions compared to previous reviews (Frich et al., 2015; Geerts et al., 2020; Husebø & Akerjordet, 2016; Rosenman et al., 2014; Sadowski et al., 2018; Steinert et al., 2012; Straus et al., 2013). In many studies, it is still not clear whether best practices for design, delivery, and evaluation are being followed (Thomas et al., 2016). It is also not clear whether there are sufficient behavioural and organisational outcomes to justify the considerable and increasing investments in medical leadership development.

Compared with previous reviews of evaluations of medical leadership development programmes for doctors, we found an increase in the proportion of studies which reported the use of active learning methods such as project work, simulation, discussions and reflections, which are widely accepted to be a vital component of leadership development (Rabin, 2014) and which were associated in our review with increased Kirkpatrick Level Four outcomes.

No single leadership development content area was particularly associated with improved outcomes. With respect to educational methods, however, there was an association between the inclusion of individual or group project work and of mentoring with organisational outcomes. This may support the established position that educational methods are more important than specific curriculum content for leadership development (for example, Rabin, 2014; West et al., 2015). Simulation and role play were less common in higher reliability studies which reported organisational outcomes than those that did not report organisational outcomes. This unexpected finding could result from these studies being situated in a training environment rather than a working environment.

Alternatively, it could result from the evaluation process and study designs rather than from a lack of organisational impact. Studies which included simulation and role play tended to focus their evaluations on objective changes in behaviour at the expense of evaluating organisational outcomes (see Table 17 in Appendix Two – Supplementary data for Chapter Ten). Interestingly, lacking a leadership development framework did not seem to impede programmes from reporting organisational outcomes. This may indicate that programmes which are designed as bespoke solutions to local needs are more likely to achieve organisational impact than pre-packaged approaches to leadership development.

There was an additional association of more senior participant level with organisational results. This may be related to the wider scope of influence or practice of senior physicians compared to resident physicians. It could also indicate that there is a longer post-programme development period before residents are able to have an impact on organisational results. This would align with our finding that programmes which evaluated longer-term outcomes were more likely to report organisational reports.

Importantly, our findings indicated that leadership development interventions which used a combination of internal and external faculty were most likely to report organisational results, and those interventions which used external faculty only were least likely. This could have important implications for procurement and design of leadership development interventions across healthcare, particularly as programmes run internally are associated with significantly reduced costs (Gagliano et al., 2010; MacPhail et al., 2015).

As in previous physician leadership development reviews that used critical appraisal instruments (Geerts et al., 2020; Rosenman et al., 2014), we found that studies frequently did not meet criteria for high reliability. Many studies failed to report important methodological features, which restricts readers' ability to appraise studies and learn

from their findings. This was particularly notable in terms of questionnaire design, with fewer than one in ten studies using validated questionnaires or reporting their questionnaire content in detail. Most studies also did not report or analyse outcome evaluation data comprehensively. Many study designs were biased towards obtaining positive results, particularly in terms of the absence of control groups, having stringent or undisclosed selection criteria, including leading questions on questionnaires, and relying solely on self-ratings. This is likely to have resulted in improved reported outcomes. The lack of evaluation quality seems to indicate perfunctory attention paid to evaluation design and precludes confident conclusions from these studies. Future studies could benefit from consulting study quality appraisal checklists such as the MERSQI and JBI in advance, in order to effectively design their evaluations.

This review does indicate that certain recommendations for improved programme evaluation are beginning to be applied into research. Whereas only 29% of the studies reviewed by Frich et al. (2015) included qualitative components, 63 (54%) of the 117 studies included in our review used mixed or qualitative methods. As outlined in Chapter Three, on a nascent and complex field such as medical leadership development research (Frich et al., 2015; Frye & Hemmer, 2012; Geerts et al., 2020; West et al., 2015), qualitative methods can have value in terms of establishing effective programme design features to achieve desired outcomes (Lockwood et al., 2015; Munn et al., 2014; Thomas et al., 2016), as well as helpful nuances of how, for whom, to what extent, or in what circumstances interventions are effective or not (Geerts et al., 2020; Kwamie et al., 2014; Steinert et al., 2012).

Additionally, many studies in this systematic review evaluated outcomes at Kirkpatrick Level Three behavioural change (57%) or Level Four organisational results (24%). This is a significant improvement from previous reviews (Frich et al., 2015; Rosenman et al., 2014; Straus et al., 2013). Changes in behaviour (Level Three) and organisational results (Level Four) are more closely associated with transfer of learning to the working

environment than participant reaction (Level One) and learning (Level Two) (Kennedy et al., 2014; Klein & Kozlowski, 2000; Saks & Burke, 2012).

Limitations and strengths

This review was limited by the reliability of the studies included. We attempted to control for study reliability using critical appraisal tools with cut-off scores for higher reliability studies.

To the best of our knowledge this is the first systematic review of medical leadership development interventions to use the JBI critical appraisal tool to critically appraise qualitative studies. The JBI tool enabled us to identify 12 additional higher reliability qualitative and mixed-methods studies which were not identified using the MERSQI. Marked heterogeneity of studies and evaluations precluded a formal meta-analysis, therefore, we adopted a mixed-methods approach. This enabled us to highlight design components that are correlated with behavioural and organisational results in higher reliability studies.

A substantial majority of studies reported only positive outcomes, which could represent a publication bias, and we limited our review to English language peer reviewed studies. In line with Frich et al. (2015), our database search was limited to Medline, however, we augmented our database search with an extensive hand-search of reference lists and citations using Web of Science and Google Scholar. The hand search revealed that many relevant empirical studies were absent from recent reviews despite some of those reviews searching a greater range of research databases. This could indicate flaws in medical leadership development literature tagging and filing procedures within medical and educational databases.

Chapter Conclusion

The systematic review presented in this chapter was performed to explore medical leadership development research regarding qualified doctors, to establish the research gap, to contextualise the evaluations presented in Part II of this thesis, and to explore common flaws in the current literature.

Additionally, this chapter has specific practical implications for those commissioning, designing and evaluating medical leadership development programmes. No specific area of curriculum content and no particular medical leadership development framework was clearly associated with behavioural or organisational outcomes. While relevance and appropriateness of educational content is important (Thomas et al., 2016), this systematic review has more clear implications for leadership development methods than for specific content. Where possible, interventions should include projects and individual or group mentoring. Transfer of learning from the programme into learners' daily work and their organisations should be planned into the programme and where possible active learning educational designs should be employed, including opportunities for learners to set their own goals for development. External faculty should be judiciously used to supplement in-house faculty, not as a replacement for in-house expertise.

In terms of evaluation design, efforts should be made to ensure that evaluations are cost-effective and produce data that is useful for both practitioners and researchers (see Chapter Nine) (Edmonstone, 2013; Patton, 2008). Effective mixed-methods evaluation strategies should be integrated into evaluation designs, as suggested in Chapter Three. Study quality checklists such as the MERSQI and JBI could be consulted in the programme design phase to help build high quality quantitative and qualitative evaluation methods into programmes. At the minimum, evaluation design should include consideration of assessment at multiple time points, inclusion of control groups and collection of objective data, as well as collection of qualitative data from interviews, focus

groups, questionnaires or observations. Programme goals and intended organisational outcomes should be explicitly considered during evaluation design (Edmonstone, 2013) so that measures of organisational outcomes (including project outcomes) can be incorporated into the evaluation design. Improving study design and building robust evaluation methods into programmes will allow evaluators and educators to more effectively understand factors which are reliably associated with high level programme outcomes. This could both inform the improvement of individual programmes and contribute to the medical leadership literature as a whole. It is only through more considered and thorough evaluation of physician leadership development programmes that we will be able to justify the investment they represent.

PART V:

Discussion

Chapter Eleven – Discussion, Limitations, Conclusions and Implications for Practice and Research

The focus for this thesis has been development of insights into evaluation methodologies for medical leadership development programmes.

In Part I, I established that there is a lack of effective evaluation methods for medical leadership development. In Part II, I established that there is a gap in medical leadership development in medical schools. In Part III, I developed, piloted and analysed a set of new medical leadership development programme evaluation questionnaires, before returning to stakeholder interviews to establish outcomes of importance. In Part IV, I conducted a systematic review of studies evaluating medical leadership development programmes for qualified doctors in order to better understand the state of current research and to focus future research.

In Part V, I will summarise the key points from each of the four preceding Parts. I will then explore limitations, conclusions and implications of my research.

Part I – Introduction, Literature Review and Methodology

In recent years, effective medical leadership has been recognised across healthcare as a vital component in the saving of money, staff turnover and patient lives. There is a growing body of research, expert opinions and political opinions that support this position (Bohmer, 2012; Darzi, 2008; Falcone & Satiani, 2008; Goodall, 2011; Spurgeon et al., 2015; Tasi et al., 2019; West et al., 2015). The importance of effective medical leadership has particularly been emphasised during the SARS-CoV-2 pandemic (for example, Lloyd-Smith, 2020; Stoller, 2020).

Increased perceived importance of medical leadership has resulted in an explosion of both leadership development programmes and investments of public money. In the United Kingdom, best estimates indicate that NHS trusts spend up to 29% of their organisational development budget on leadership development (West et al., 2015), which could amount to as much as £1 billion per year. This trend is limited neither to the United Kingdom nor healthcare. Beer et al. (2015) estimate that worldwide, more than $350 billion US is spent on leadership development in American companies.

Such expenditure relies on a key assumption and conviction that leadership development programmes lead to significant clinical impact. The lack of effective evaluation of these programmes leaves this an untested assumption and was the focus of my doctoral research at Oxford.

As outlined in Chapter Two, quality of evaluations of medical leadership development programmes has been limited by lack of consensus regarding the meaning of leadership and the intended outcomes of medical leadership development programmes. Historically, there have been widely varied academic definitions of leadership (Northouse, 2013; Parry & Bryman, 2013; Stogdill, 1950). There are also many lay definitions of leadership. This ambiguity allows for misalignment of expectations for programmes between commissioners, providers, participants and researchers.

As explored in Chapter Three, there are few established and validated methods for evaluation of medical leadership development programmes (Frich et al., 2015; Geerts et al., 2020; Lyons et al., 2020) and for evaluation of leadership development in general (Hannum et al., 2007). Consequently, providers, programme commissioners and researchers have developed bespoke evaluation tools and methods for each programme evaluation. These bespoke evaluations are frequently developed entirely from scratch or from previous flawed evaluations and often do not meet expected standards for high-

quality research, despite clear standards existing (Cook, 2010; Hannum, 2004; Patton, 2013; Reed et al., 2007, 2008; Rosenman et al., 2015; Sullivan, 2011).

Part II – Leadership development in medical schools

In Part II, I first conducted a needs analysis of leadership development in medical schools in Australia and New Zealand, which I have presented in Chapter Four. In this needs analysis, published in BMJ Leader (Lyons et al., 2019), I showed that while current standards provide clear opportunities for medical students to develop foundational competency in 60% of the Medical Leadership Competency Framework (MLCF) elements, there remain important curriculum gaps in the Australian Medical Council's required Outcomes for Graduates.

These Outcomes for Graduates standards form the basic requirements for a medical degree in Australia and New Zealand, and medical schools have scope to stretch beyond these standards to provide more complete learning experiences for their medical students.

To investigate where medical schools exceed these standards, I conducted a systematic review of medical leadership development programmes in medical schools internationally (Lyons et al., 2018), which I have presented in Chapter Five. In this chapter, which was published in the New Zealand Medical Journal, I found only eleven studies of leadership development programmes in medical schools that evaluated their programmes beyond learner reactions. I noted generally positive outcomes from those studies regardless of their time cost and I noted that there is need for more work on development of evaluations and evaluation methods.

Part III – Development of evaluation methods for medical leadership development programmes

In Part III, with a view to producing research with potential for impact, I moved the focus of my research from leadership development for medical students through to leadership development for practicing doctors. I focused on adapting existing medical leadership frameworks into evaluation tools and on refining and improving these tools through mixed-methods programme evaluations. For this work on evaluation development, I used case studies as presented in Chapter Three and in Appendix Six – Case study descriptions.

Chapter Six

In Chapter Six, I developed a new summative leadership development questionnaire from an existing formative questionnaire and piloted this questionnaire in Emerging Leaders Programme Cohort One (ELP1). The results from this pilot questionnaire showed a statistically significant increase in participants' self-assessed knowledge of their organisation's strategy. Despite taking care in the design of this questionnaire, however, basing it on an existing model and conducting an expert review of the questionnaire, I was unable to critique the validity of the findings. Single-item questions meant that I was unable to conduct exploratory factor analyses to check for underlying constructs, and increased the vulnerability of question items to response-shift bias; lack of objective measures of change to correlate with each of the items meant that I was unable to assess the concordance with objective measures (criterion validity); because I did not have a suitable existing questionnaire to co-administer it was not possible to assess convergent validity; and because there were no decisions made as a result of the Primary Colours score for individuals or for the programme, consequential validity was not relevant (Cook et al., 2015; Kane, 2001; Maul, 2018; Shaw & Crisp, 2015; Sullivan, 2011).

The validity argument (Cook et al., 2015; Kane, 2001) for the Primary Colours Questionnaire would be that an increase in score from pre- to post-programme self-assessments means that there has been a corresponding increase in a participant's capacity in each of the associated Primary Colours domains. Significant further work would be required for this to be a credible argument, given that I did not collect appropriate data to investigate reliability, criterion, convergent or consequential validity.

It is worth pointing out that by simply basing questionnaire development on an existing model and conducting an expert review, the Primary Colours Questionnaire has had more investigation of validity than nine in ten papers as reviewed in Chapter Ten. Rather than a justification of the limitations of the Primary Colours Questionnaire, this is an indictment of the lack of attention to validation in the research literature surrounding medical leadership development programmes.

It is vital for an evaluation to produce credible data and make valid inferences if researchers and stakeholders are to be able to make decisions based on the evaluation of a programme. Further studies would be needed for the Primary Colours Questionnaire to reach this standard. In the meantime, the Primary Colours Questionnaire produced academically interesting findings, and provided a foundation for further evaluation methods development in Chapter Seven and Chapter Eight.

Chapter Seven

In Chapter Seven, I explored the effect of response-shift bias on the Primary Colours Questionnaire by applying the questionnaire as both a prospective pre- and post-programme test and as a retrospective then- and post- programme test. I also co-administered an adapted version of the MLCF self-assessment questionnaire. To my knowledge there are no published examples of the MLCF self-assessment being used as an evaluation tool, though the MLCF domains and competencies have been evaluated (Cohen et al., 2017). Chapter Seven demonstrated that retrospective application of the

Primary Colours Questionnaire resulted in a greater number of statistically significant changes in questionnaire item responses. This suggested that the new questionnaire could be vulnerable to response-shift bias and highlighted the potential of retrospective self-assessment as a means to detect response-shift bias.

The evaluation presented in Chapter Seven (of ELP2) was more extensive than the pilot presented in Chapter Six (of ELP1). I still, however, did not have reliable objective indicators of changes in knowledge and behaviour which I could correlate with changes on the Primary Colours Questionnaire. This prevented me from evaluating the validity argument for the use of retrospective then-tests.

There were, nonetheless, some interesting findings which arose from co-administration of the Primary Colours Questionnaire with the MLCF self-assessment questionnaire. The MLCF self-assessment questionnaire relies on behavioural questions with less participant interpretation required than for the Primary Colours Questionnaire. Importantly, the comparison between retrospective and prospective changes in scores showed that the MLCF scores were less affected by response shift bias. This is consistent with research outside medical leadership development, which indicates that behavioural questions tend to be less vulnerable to response-shift bias than abstract or attitudinal questions (Hill, 2020; Hill & Betz, 2005).

The MLCF self-assessment is a moderately large questionnaire, with 56 questions. If changes in the Primary Colours Questionnaire offer a good predictive relationship to the MLCF, it would be valuable to have a shorter and more easily administered questionnaire than the MLCF self-assessment. The outcomes of interest to decision-makers (be they funders, faculty or prospective participants) are, however, unlikely to be changes in a questionnaire score. It would therefore be important to assess the validity argument of the MLCF questionnaire also, with respect to outcomes of interest to decision-makers.

This would be a worthwhile future research area, given that the MLCF has been widely accepted and has recently been adopted into GMC standards (Cohen et al., 2017).

Chapter Eight

In Chapter Eight, I developed a Supplementary Questionnaire in collaboration with a participant on the Future Leaders Programme (FLP 2016-2017). This Supplementary Questionnaire explored both the impact of a programme on a participant in their life outside the programme and the impact on their wellbeing. Questionnaire outcomes across all four of the case studies it was applied in (FLP 2016-2017, ELP1, ELP2, ELP3) indicated a positive general impact on wellbeing and a high self-reported application of learning from the programme into day to day life. Further, there was concordance between quantitative questions regarding the impact of a programme on wellbeing and the free-text comments from participants. In order to assess whether it would be worthwhile conducting further studies to assess the convergent validity of the Supplementary Questionnaire with existing validated measures for wellbeing, it would be important to evaluate how important these wellbeing outcomes are to organisations, programmes and healthcare professionals. This conclusion, which was mirrored in Chapters Six and Seven, led to my work in Chapter Nine, which explored the intended outcomes of leadership development programmes, and the uses of evaluation.

Chapter Nine

Given the lack of established definitions of medical leadership and established goals of medical leadership development programmes, it is unsurprising that evaluations of medical leadership development programmes have tended to neglect evaluation of outcomes at behavioural and results levels (Frich et al., 2015; Geerts et al., 2020; Kumar et al., 2020; Lyons et al., 2018, 2020; Sultan et al., 2019). For many (if not most) programmes, it is not clear what effective outcomes would constitute in terms of behavioural change or organisational results.

Traditional training programmes have often considered success to mean financial return-on-investment (Avolio et al., 2010). In a medical leadership development programme this would mean evaluating whether savings from improvements in healthcare quality would exceed the cost of the leadership programme (Hannum et al., 2007).

Development programmes usually have a range of outcomes beyond financial return-on-investment, including participant reaction, learning, behavioural change and tangible results (Kirkpatrick, 1979). An effective utilisation-focused evaluation would explore what a successful outcome would mean for each stakeholder group at each level of outcome (Grove et al., 2005; Kirkpatrick, 1979; Patton, 2008), and plan evaluation methods which would explore whether these outcomes were obtained.

Chapter Nine was conducted to explore variation in expected use of evaluations and expected outcomes for programmes from the perspectives of key stakeholders. I was limited in my ability to collect data for this chapter because of the Covid-19 pandemic, because my planned interviewees all worked in healthcare. Nevertheless, through analysis of the interviews I was able to carry out, combined with data from other related interviews, Chapter Nine suggests that there is considerable breadth in outcomes which key stakeholders consider to be important for an effective leadership development programme. Further investigation is warranted into how evaluation can be tailored to consistently provide data which supports both decision-making by these key stakeholders and development of research knowledge.

Findings from the exploratory research in Chapter Nine further suggest that the breadth of programme outcomes expected by key stakeholders are not encapsulated in the medical leadership competency models that I have explored over the course of this thesis (the MLCF and the Primary Colours Model). This is not surprising, given that neither were primarily developed as evaluation tools but rather as descriptive and guiding

models. As a result, however, the tools I had been refining from those frameworks will be helpful but not sufficient for summative or formative evaluation. Overall, it appears that we need to acquire a more complete understanding of what programmes are being developed to achieve, in order to develop evaluation methods which can inform programme improvement and justify ongoing expenditure on medical leadership development.

Part IV - Systematic Review of Medical Leadership Development Programmes for Qualified Doctors

Chapter Ten

In Chapter Ten, which was published in BMJ Leader in 2020 (Lyons et al., 2020), I contextualised the research I presented in Part III by systematically reviewing and critically appraising the existing literature on medical leadership development programmes for qualified doctors.

In this second systematic review, I had a particular focus on critically appraising the quality of evaluations. There had been variable quality of studies in the systematic review presented in Chapter Five (Lyons et al., 2018), which was supported by other systematic reviews (Frich et al., 2015; Geerts et al., 2020; Kumar et al., 2020; Rosenman et al., 2015; Sultan et al., 2019).

It was disappointing to see that more than three quarters of the included evaluations (89 out of 117) failed to meet even half the established criteria for higher-quality studies. Two common and important causes for the lack of high quality research in medical leadership development were again indicated: lack of established evaluation methods and lack of clarity regarding intended programme outcomes (Frich et al., 2015; Geerts et al., 2020; Rosenman et al., 2014; Straus et al., 2013).

Discussion

While there are links between effective medical leadership and healthcare outcomes (Clay-Williams et al., 2017; Goodall, 2011), many medical leadership development programmes only consider outcomes related only to reaction, learning or changes in behaviour, and fail to evaluate the organisational or patient impact of their programmes (Frich et al., 2015; Geerts et al., 2020; Lyons et al., 2020). There are a range of evaluation methods and methodologies that have been used to evaluate leadership development programmes, and many of these could be easily adapted for use in healthcare (Hannum, 2004; Patton, 2008) (see Chapter Three).

Part of the reason for the lack of higher-level evaluations may be the lack of investment into medical education research in general and medical leadership development research specifically (Reed et al., 2007). More extensive evaluations require more extensive investments of time and money (Hannum et al., 2007; Kennedy et al., 2014; Patton, 2008). Reed's research into the association between medical research quality and the funding available showed that 71% of medical education research is completed without any funding (Reed et al., 2007). Kennedy and colleagues found that lack of resources including time and budget were the main reason why training professionals omitted evaluation at the behavioural and organisational levels (Kennedy et al., 2014).

In the face of tight budgets and performance indicators, it is easy to see why evaluation might be given low priority. There is, however, a range of research which indicates effective evaluation is associated with more effective programme outcomes (Latham & Steele, 1983; Saks & Burke, 2012; Watkins et al., 2011).

Davidson and Martineau suggest that as a rule of thumb, evaluation should be allocated 10% of the funding for any given programme (Davidson & Martineau, 2007). Given that the NHS Leadership Academy alone spent £50 million in 2015/2016 and £70

million in 2014/2015 on leadership programmes (NHS Improving Quality, 2016), Davidson and Martineau's rule of thumb would indicate that the NHS Leadership Academy alone should be spending at least £5 million per year on evaluations. I could not find any evidence to suggest this is the case.

Davidson and Martineau also point out that if programmes are not evaluated effectively, "funds could be poured into a program that does not work" (Davidson & Martineau, 2007, p. 457). Conversely, and perhaps even more importantly for organisations that promote the development of medical leadership, they emphasise that insufficient investment into evaluation prevents key stakeholders from learning about the outcomes of a programme and the value of those outcomes to an organisation: "the program may have been quite valuable in producing strategically relevant outcomes, but the sponsors do not know how to replicate the outcomes in future initiatives and do not have the data they need to illustrate the strategic value of the initiative to their own stakeholders" (Davidson & Martineau, 2007, p. 457).

In addition to lack of time and budget allocated to evaluation of higher-level outcomes, Kennedy and colleagues found that institutional factors such as perceived low importance of evaluation by organisations and departments, lack of support from organisational management and lack of expertise in evaluation were critical factors leading to non-evaluation of higher-level outcomes (Kennedy et al., 2014). They also found that support from organisational management for conducting an evaluation was the top factor that facilitated evaluation of behavioural and organisational outcomes from training programmes.

The existing focus in modern healthcare on evidence-based practice could provide an excellent position from which to systematically support more effective evaluation of outcomes from leadership development programmes. In addition to helping us understand the effectiveness of investments in leadership development, supporting

evaluation from an organisational level will facilitate the development of new knowledge around optimal medical leadership development. New knowledge around optimal medical leadership development could have considerable financial implications, organisational implications, and could improve the quality of the care we provide.

In the early stages of this research, I was surprised to find how much public money is spent on leadership development both broadly and in healthcare specifically. As quoted above, perhaps £70 million is spent per year by the NHS Leadership Academy alone (NHS Improving Quality, 2016), and as much as £1 billion may be spent across the NHS (West et al., 2015). Yet studies still suggest that at least half of doctors in training feel that their leadership training is inadequate, and perhaps only a third of medical schools have leadership development fully integrated into their curriculum (Baird et al., 2018; Jefferies et al., 2016; Neeley et al., 2017).

The tendency towards low quality evaluations that focused on participant reactions as a primary outcome of a medical leadership development programmes was also surprising. This was seen in my systematic reviews of medical leadership development in medical schools (Chapter Five) (Lyons et al., 2018) and in qualified doctors (Chapter Ten) (Lyons et al., 2020). Low quality and reaction-focused evaluations have also been highlighted in a range of other systematic reviews (for example, Frich et al., 2015; Geerts et al., 2020; Kumar et al., 2020; Rosenman et al., 2014; Sultan et al., 2019; Webb et al., 2014). Chapter Three (Methodology) summarises a range of research that indicates that evaluation of medical leadership development is not easy. Nonetheless, if medical leadership development programmes are important and they are to be integrated into already cramped curricula (Jefferies et al., 2016), higher quality evaluations that focus on more tangible impacts will be vital.

Other important unexpected findings from this research included apparent decrease in participants' self-assessments after FLP 2016-2017, which led to the research into

response shift bias and retrospective self-assessments (see Chapter Seven); participants noting that the existing evaluation of FLP 2016-2017 was missing outcomes that were important to them, including the impact on their wellbeing, which led to development of the Supplementary Questionnaire (see Chapter Eight); and the variation in expectations from different stakeholders for programmes and for programme evaluations that became clear through early and informal interviews and was briefly explored in Chapter Nine.

Limitations

As evidenced by the systematic reviews in Chapters Five and Nine, evaluation of medical leadership development is still nascent, and there is substantial room for improvement.

Definition

There is dissent over the most appropriate definitions for leadership, let alone for medical leadership specifically, and therefore I adopted a broad definition that can provide insights to a wide range of applications (Yukl, 2013): *"*achieving results with and through others*"* (Blake & Mouton, 1985, p. 198). This broad definition still presents a limitation for my research, as there will inevitably be dissent as to whether this is an appropriate way to define medical leadership. As outlined in Chapter Three, the lack of an established definition for medical leadership causes an ontological and epistemological tension for this research. Rather than focusing on these broad questions, I have elected to take a pragmatic approach to my research, informed by Michael Quinn Patton's principles of utilisation-focused evaluation.

Unclear stakeholder expectations

Programme goals are frequently not explored extensively enough to properly inform evaluation (Erickson & Noonan, 2018; Patton, 2008). While my research in this thesis

was grounded in utilisation-focused evaluation, it became clear to me through informal conversations and through analysis of the results of questionnaire development (Part III) that there is still considerable additional room for clarification of the expectations of users of evaluations. In Chapter Nine, I began the work of exploring the range of uses and intentions for leadership development evaluations in healthcare. While this research phase was cut short by the pandemic, it suggested that there is clear room for future research into the expectations and needs of stakeholders and decision-makers regarding medical leadership development programmes.

Lack of validation

In addition to further clarifying the outcomes that are important for decision-makers, it is important to establish evaluation methods which can provide credible and valid data regarding those important outcomes. Validation of questionnaires was not the focus of this thesis. My research nonetheless suggests that there is considerable room for future research into validation of medical leadership development programme evaluation methods. Disagreements regarding validation have raged since the 1970s, and as a result validation remains a black-box for many researchers (for example, D. A. Cook et al., 2015; Kane, 2001). While guides and primers have been published for medical education validation studies (Cook et al., 2015; Sullivan, 2011), there is room for much more clarification and simplification, so as to make validation studies more approachable for (mostly unfunded) researchers (Reed et al., 2007). In the questionnaires I developed in this thesis, evaluating validity arguments with a focus on outcome data that will assist decision-makers (including funders, curriculum leads, faculty and participants), is important if these are to be used in future programmes.

Adaptation of this research

Considerable gaps remain in medical leadership development research. As I progressed through the research I have presented in this thesis, I uncovered further

assumptions within both the literature and within my research question. As a result, my research has remained broad and exploratory, with each part built on the research in earlier parts. While this meant that I was not able to predict the direction of my research at the beginning, it has been a strength. I have been able to adapt and adjust my research to focus on investigating these assumptions and to solidifying the foundation for further research into medical leadership development.

Research Implications

I have separated some of the key implications of this research into practical implications for evaluation and implications for future research.

Practical implications for evaluation of leadership development

1. Evaluations should be integrated into leadership development programmes from the design phase.
2. During the design phase, clear explicit and implicit objectives should be specified with the key stakeholders. These objectives for the programme should be revisited at fixed time points to ensure that a programme remains aligned with stakeholder objectives.
3. Evaluators and providers should consult the primary intended users of evaluations and clarify:
    a. How those users will use the evaluation data; and
    b. What would constitute credible and valid data for the purpose of their decisions.
4. Established checklists for high-quality evaluations such as the MERSQI and the JBI should be consulted with a view to improving the quality of research, and evaluators should consider collection of data regarding outcomes at each of Kirkpatrick's levels.

5. Evaluators should incorporate multiple methods into their evaluations where possible and use qualitative methods to allow for unexpected data.
6. Evaluators should develop a plan for longitudinal evaluation of programmes where feasible, in order to capture long-term as well as short-term outcomes.
7. Where possible, evaluators should look for appropriate control groups which can help to elucidate environmental effects, maturation and other confounders.
8. Where there is potential for response-shift bias, retrospective then-tests can be used to help identify this bias when applied in parallel with prospective pre-tests.
    a. Further research is needed to establish the appropriateness of retrospective then-post tests in isolation, as there may be variable effects and other biases that need to be explored.

Areas for further research

1. There is a need for further exploration of common intended outcomes and uses of evaluation from key stakeholder groups in medical leadership development. A larger-scale research project would be of value here, as there are significant implications for medical leadership development evaluations and for development of the literature.
2. Once there is greater understanding of (1), it would be helpful to map existing evaluation methods and methodologies with specific common outcomes of interest and evaluation uses, so that inexperienced evaluators have a firm starting point for evaluations.
3. The mapping in (2) will highlight areas where there is need for more development of evaluation methods. It is likely that there are appropriate methods in other disciplines which can be applied to medical leadership development evaluations effectively.

4. There is considerable need for further research into medical leadership development. It is therefore important that educators and researchers communicate their evaluation findings. By mapping out important evaluation methods to important and common outcomes and uses, it may be possible to facilitate more publication of research regarding medical leadership development programmes.
5. There is limited understanding of the relationship between programme design and outcomes. It may be possible for inventive and ambitious researchers to develop study designs to test the impact of various design features. Where this is not feasible, there is space to explore possible relationships between programme designs and outcomes, in order to establish good practice.

Personal Reflection

Early in my doctoral research, I realised my plan to identify best practices in medical leadership development relied on an incorrect assumption that there would be established methods for evaluating the successfulness of a medical leadership development programme. To address this underlying assumption, I adapted my research question to focus directly on evaluation of medical leadership development programmes. Later in my research, after analysing and critically appraising the results of the questionnaires I had developed, I realised that the needs of decision-makers regarding the programmes I was evaluating often remained obscure, and I again needed to adapt my plan, to include research into these needs.

Reflecting on my development as a researcher, it is clear to me that I now have more tools in my repertoire for recognising, understanding and addressing fundamental assumptions. As I look towards further work improving medical leadership development, I find myself seeing such assumptions both as significant barriers to progress and important areas to investigate further.

Large investments continue to be made into medical leadership development, with minimal evidence that these achieve any impact beyond happy learners. There appears to be an underlying assumption held by commissioners that investing in medical leadership development will inevitably improve medical leadership and thereby effect positive impact on healthcare outcomes. This assumption remains for the most part untested.

If we want to ensure that investments into medical leadership development programmes achieve the ends that commissioners, organisations and politicians intend, we need to get better at questioning our assumptions and evaluating our investments into medical leadership development programmes.

References

Abbas, M. R., Quince, T. A., Wood, D. F., & Benson, J. A. (2011). Attitudes of medical students to medical leadership and management: A systematic review to inform curriculum development. *BMC Medical Education*, *11*(1), 93. https://doi.org/10.1186/1472-6920-11-93

ACT Academy. (2012). *Stakeholder Analysis*. https://doi.org/10.1201/b12133-8

Aggarwal, R., & Swanwick, T. (2015). Clinical leadership development in postgraduate medical education and training: Policy, strategy, and delivery in the UK national health service. In *Journal of Healthcare Leadership* (Vol. 7, pp. 109–122). Dove Medical Press Ltd. https://doi.org/10.2147/JHL.S69330

Agius, S. J., Brockbank, A., Baron, R., Farook, S., & Hayden, J. (2015). The impact of an integrated medical leadership programme. *Journal of Health, Organisation and Management*, *29*(1), 39–54. https://doi.org/10.1108/JHOM-09-2013-0188

Al-Mutawa, N., Elmahdi, H., & Joyce, P. (2016). The implementation of a practice management programme for family medicine residents in Qatar. *Education for Primary Care*, *27*(5), 380–385. https://doi.org/10.1080/14739879.2016.1219620

Allen, J. M., & Nimon, K. (2007). Retrospective pretest: A practical technique for professional development evaluation. *Journal of Industrial Teacher Education*, *44*(3), 27–42.

Andersen, J. A. (2006). Leadership, personality and effectiveness. *Journal of Socio-Economics*, *35*(6), 1078–1091. https://doi.org/10.1016/j.socec.2005.11.066

Ardichvili, A., Natt och Dag, K., & Manderscheid, S. (2016). Leadership Development: Current and Emerging Models and Practices. *Advances in Developing Human*

Resources, *18*(3), 275–285. https://doi.org/10.1177/1523422316645506

Aromataris, E., & Munn, Z. (2019). *JBI Reviewer's Manual*. Joanna Briggs Institute. https://doi.org/10.46658/JBIRM-19-01

Arthur, W., Bennett, W., Edens, P. S., & Bell, S. T. (2003). Effectiveness of training in organizations: A meta-analysis of design and evaluation features. *Journal of Applied Psychology*, *88*(2), 234–245. https://doi.org/10.1037/0021-9010.88.2.234

Australian Medical Council. (2012). *Standards for Assessment and Accreditation of Primary Medical Programs by the Australian Medical Council 2012*. http://www.amc.org.au/files/d0ffcecda9608cf49c66c93a79a4ad549638bea0_original.pdf

Avolio, B. J., Avey, J. B., & Quisenberry, D. (2010). Estimating return on leadership development investment. *Leadership Quarterly*, *21*(4), 633–644. https://doi.org/10.1016/j.leaqua.2010.06.006

Avolio, B. J., Reichard, R. J., Hannah, S. T., Walumbwa, F. O., & Chan, A. (2009). *A meta-analytic review of leadership impact research: Experimental and quasi-experimental studies*. https://doi.org/10.1016/j.leaqua.2009.06.006

Avolio, B. J., Walumbwa, F. O., & Weber, T. J. (2009). Leadership: Current theories, research, and future directions. *Annual Review of Psychology*, *60*(1), 421–449. https://doi.org/10.1146/annurev.psych.60.110707.163621

Baird, H., Malik, M., Patel, N., Yassaee, A. A., Tweedie, J., & Hynes, G. (2018). Supporting junior doctors' leadership: a trilogy of work. *BMJ Leader*, *2*(4), 126–127. https://doi.org/10.1136/leader-2018-000125

Barge, S., & Gehlbach, H. (2012). Using the Theory of Satisficing to Evaluate the Quality of Survey Data. *Research in Higher Education*, *53*(2), 182–200.

https://doi.org/10.1007/s11162-011-9251-2

Barr, H. (1999). Evaluating Interprofessional Education: two systematic reviews for health and social care. In *British Educational Research Journal* (Vol. 25, Issue 4). https://doi.org/10.1080/0141192990250408

Bass, B. M. (1985). *Leadership and performance beyond expectations*. Free Press ; Collier Macmillan.

Bates, R. (2004). A critical analysis of evaluation practice: The Kirkpatrick model and the principle of beneficence. *Evaluation and Program Planning*, *27*(3), 341–347. https://doi.org/10.1016/j.evalprogplan.2004.04.011

Beckman, T. J., & Cook, D. A. (2007). Developing scholarly projects in education: A primer for medical teachers. *Medical Teacher*, *29*(2–3), 210–218. https://doi.org/10.1080/01421590701291469

Beer, M., Finnström, M., Schrader, D., & Zank, B. (2015). Why Leadership Training Fails—and What to Do About It. *Harvard Business Review*, *2016*(October), 50–58.

Bennett, N., Wise, C., Woods, P., & Harvey, J. A. (2003). Distributed leadership: A review of literature carried out for NCSL. In *National College for School Leadership*. http://www.ncsl.org.uk

Bennis, W. G., & Nanus, B. (1986). *Leaders : the strategies for taking charge*. Perennial Library.

Berghout, M. A., Fabbricotti, I. N., Buljac-Samardžić, M., & Hilders, C. G. J. M. (2017). Medical leaders or masters?—A systematic review of medical leadership in hospital settings. *PLOS ONE*, *12*(9), e0184522. https://doi.org/10.1371/journal.pone.0184522

Bergman, D., Savage, C., Wahlstrom, R., & Sandahl, C. (2008). Teaching group dynamics - Do we know what we are doing? An approach to evaluation. *Medical Teacher*, *30*(1), 55–61. https://doi.org/10.1080/01421590701758624

Berkenbosch, L., Muijtjens, A. M. M., Zimmermann, L. J. I., Heyligers, I. C., Scherpbier, A. J. J. A., & Busari, J. O. (2014). A pilot study of a practice management training module for medical residents. *BMC Medical Education*, *14*(101088679 PG-107), 107. https://doi.org/10.1186/1472-6920-14-107

Bhanji, F., Donoghue, A. J., Wolff, M. S., Flores, G. E., Halamek, L. P., Berman, J. M., Sinz, E. H., & Cheng, A. (2015). Part 14: Education: 2015 American Heart Association guidelines update for cardiopulmonary resuscitation and emergency cardiovascular care. *Circulation*, *132*(18), S561–S573. https://doi.org/10.1161/CIR.0000000000000268

Bhanji, F., Gottesman, R., De Grave, W., Steinert, Y., & Winer, L. R. (2012). The retrospective pre-post: A practical method to evaluate learning from an educational program. *Academic Emergency Medicine*, *19*(2), 189–194. https://doi.org/10.1111/j.1553-2712.2011.01270.x

Bhatia, K., Morris, C. A., Wright, S. C., Takayesu, J. K., Sharma, R., & Katz, J. T. (2015). Leadership Training for Residents: a Novel Approach. *Physician Leadership Journal*, *2*(2), 76–80.

Bickley, L. S., Bates, B., & Szilagyi, P. G. (2013). *Bates' pocket guide to physical examination and history taking* (7th ed.). Wolters Kluwer/Lippincott Williams & Wilkins.

Biesta, G. (2010). Pragmatism and the Philosophical Foundations of Mixed Methods Research. In A. Tashakkori & C. Teddlie (Eds.), *SAGE Handbook of Mixed Methods in Social & Behavioral Research*. SAGE Publications Ltd.

https://doi.org/10.4135/9781506335193

Black, A. M., & Earnest, G. W. (2009). Measuring the Outcomes of Leadership Development Programs. *Journal of Leadership & Organizational Studies*, *16*(2), 184–196. https://doi.org/10.1177/1548051809339193

Blake, R. R., & Mouton, J. S. (1985). *The managerial grid III : a new look at the classic that has boosted productivity and profits for thousands of corporations world-wide*. Gulf Pub. Co., Book Division.

Bohmer, R. (2012). The instrumental value of medical leadership: Engaging doctors in improving services. In *The King's Fund, London*.

Bornstein, M. H., Putnick, D. L., Costlow, K. M., & Suwalsky, J. T. D. (2020). Retrospective report revisited: Long-term recall in European American mothers moderated by developmental domain, child age, person, and metric of agreement. *Applied Developmental Science*, *24*(3), 242–262. https://doi.org/10.1080/10888691.2018.1462090

Bowen, G. A. (2009). Document analysis as a qualitative research method. *Qualitative Research Journal*, *9*(2), 27–40. https://doi.org/10.3316/QRJ0902027

Boyle, D. K., & Kochinda, C. (2004). Enhancing Collaborative Communication of Nurse and Physician Leadership in Two Intensive Care Units. *Journal of Nursing Administration*, *34*(2), 60–70. https://doi.org/10.1097/00005110-200402000-00003

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Bromiley, M. (2015). The husband's story: From tragedy to learning and action. *BMJ Quality and Safety*, *24*(7), 425–427. https://doi.org/10.1136/bmjqs-2015-004129

Brown, M. E., & Gioia, D. A. (2002). Making things click. Distributive leadership in an online division of an offline organization. *Leadership Quarterly*, *13*(4), 397–419. https://doi.org/10.1016/S1048-9843(02)00123-6

Burns, J. M. (1978). *Leadership*. Harper & Row.

Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*.

Canadian College of Health Leaders. (2010). *Key Points to Leadership Growth: A Checklist for Leaders*. https://leadscanada.net/document/2570/KEYPOINTS_2019_EN.pdf

Canter, R. (2015). *New Consultants Leadership Programme Oxford University Hospitals*.

Canter, R. (2016). *Implications of Future Leaders Programme*.

Carufel-Wert, D. A., Younkin, S., Foertsch, J., Eisenberg, T., Haq, C. L., Crouse, B. J., & Frey, J. J. (2007). LOCUS: Immunizing medical students against the loss of professional values. *Family Medicine*, *39*(5), 320–325.

Clark, J., & Armit, K. (2008). Attainment of competency in management and leadership: No longer an optional extra for doctors. *Clinical Governance*, *13*(1), 35–42. https://doi.org/10.1108/14777270810850616

Clay-Williams, R., Ludlow, K., Testa, L., Li, Z., & Braithwaite, J. (2017). Medical leadership, a systematic narrative review: Do hospitals and healthcare organisations perform better when led by doctors? *BMJ Open*, *7*(9). https://doi.org/10.1136/bmjopen-2016-014474

Cohen, D., Vlaev, I., McMahon, L., Harvey, S., Mitchell, A., Borovoi, L., & Darzi, A. (2017). The Crucible simulation: Behavioral simulation improves clinical leadership

skills and understanding of complex health policy change. *Health Care Management Review*, *44*(3), 246–255. https://doi.org/10.1097/HMR.0000000000000162

Cole, D. C., Giordano, C. R., Vasilopoulos, T., & Fahy, B. G. (2017). Resident Physicians Improve Nontechnical Skills When on Operating Room Management and Leadership Rotation. *Anesthesia and Analgesia*, *124*(1), 300–307. https://doi.org/10.1213/ANE.0000000000001687

Coleman, M. M., Blatt, B., & Greenberg, L. (2012). Preparing students to be academicians: A national student-led summer program in teaching, leadership, scholarship, and academic medical career-building. *Academic Medicine*, *87*(12), 1734–1741. https://doi.org/10.1097/ACM.0b013e318271cfd6

Cook, D. A. (2010). Twelve tips for evaluating educational programs. *Medical Teacher*, *32*(4), 296–301. https://doi.org/10.3109/01421590903480121

Cook, D. A., Bowen, J. L., Gerrity, M. S., Kalet, A. L., Kogan, J. R., Spickard, A., & Wayne, D. B. (2008). Proposed standards for medical education submissions to the Journal of General Internal Medicine. *Journal of General Internal Medicine*, *23*(7), 908–913. https://doi.org/10.1007/s11606-008-0676-z

Cook, D. A., Brydges, R., Ginsburg, S., & Hatala, R. (2015). A contemporary approach to validity arguments: A practical guide to Kane's framework. *Medical Education*, *49*(6), 560–575. https://doi.org/10.1111/medu.12678

The Hawthorne Effect in Educational Research, 44 The Phi Delta Kappan 116 (2015). https://doi.org/10.2307/20342865

Cooper, S. (2001). Developing leaders for advanced life support: evaluation of a training programme. *Resuscitation*, *49*(1), 33–38. https://doi.org/10.1016/S0300-

9572(00)00345-2

Côté, L., & Turgeon, J. (2005). Appraising qualitative research articles in medicine and medical education. *Medical Teacher*, *27*(1), 71–75. https://doi.org/10.1080/01421590400016308

Cox, E. P. (1980). The Optimal Number of Response Alternatives for a Scale: A Review. *Journal of Marketing Research*, *17*(4), 407. https://doi.org/10.2307/3150495

Dalakoura, A. (2010). Differentiating leader and leadership development: A collective framework for leadership development. *Journal of Management Development*, *29*(5), 432–441. https://doi.org/10.1108/02621711011039204

Daniel, S. (2011). Satisficing in Survey Design. *Deakin Symposium on Contemporary Approaches to Research in Mathematics, Science, Health and Environmental Education*.

Dannels, S. A., Yamagata, H., McDade, S. A., Chuang, Y. C., Gleason, K. A., McLaughlin, J. M., Richman, R. C., & Morahan, P. S. (2008). Evaluating a leadership program: A comparative, longitudinal study to assess the impact of the Executive Leadership in Academic Medicine (ELAM) program for women. *Academic Medicine*, *83*(5), 488–495. https://doi.org/10.1097/ACM.0b013e31816be551

Darzi, A. (2008). *High quality care for all: NHS next stage review final report*. The Stationery Office.

Davidson, E. J., & Martineau, J. W. (2007). Strategic uses of evaluation. *The Handbook of Leadership Development Evaluation*, 433–463.

Day, D. V. (2000). Leadership development: A review in context. *Leadership Quarterly*, *11*(4), 581–613. https://doi.org/10.1016/s1048-9843(00)00061-8

Day, D. V., Fleenor, J. W., Atwater, L. E., Sturm, R. E., & McKee, R. A. (2014). Advances in leader and leadership development: A review of 25 years of research and theory. *Leadership Quarterly*, *25*(1), 63–82. https://doi.org/10.1016/j.leaqua.2013.11.004

Day, D. V., Sin, H. P., Day, V. D., & Sin, H. P. (2011). Longitudinal tests of an integrative model of leader development: Charting and understanding developmental trajectories. *Leadership Quarterly*, *22*(3), 545–560. https://doi.org/10.1016/j.leaqua.2011.04.011

Day, D. V. (2014). *Leadership: A Critical Historical Analysis of the Influence of Leader Traits*. https://doi.org/10.4324/9781315820972

Den Hartog, D. N., Van Muijen, J. J., & Koopman, P. L. (1997). Transactional versus transformational leadership: An analysis of the MLQ. *Journal of Occupational and Organizational Psychology*, *70*(1), 19–34. https://doi.org/10.1111/j.2044-8325.1997.tb00628.x

Denscombe, M. (2010). *The good research guide : for small-scale social research projects* (Fourth). McGraw Hill/Open University Press.

Department of Health. (2000). *The NHS Plan - A plan for reform, A plan for investment*. The Stationery Office.

Dillman, D. A. (2007). Mail and internet surveys: The tailored design method, 2nd ed. In *Mail and internet surveys: The tailored design method, 2nd ed.* John Wiley & Sons Inc.

Dinh, J. E., Lord, R. G., Gardner, W. L., Meuser, J. D., Liden, R. C., & Hu, J. (2014). Leadership theory and research in the new millennium: Current theoretical trends and changing perspectives. *Leadership Quarterly*, *25*(1), 36–62. https://doi.org/10.1016/j.leaqua.2013.11.005

Donaldson, S. I., Patton, M. Q., Fetterman, D. M., & Scriven, M. (2010). The 2009 Claremont Debates: The Promise and Pitfalls of Utilization-Focused and Empowerment Evaluation. *Journal of MultiDisciplinary Evaluation*, *6*(14), 159–165.

Dopson, S., Ferlie, E., McGivern, G., Fischer, M. D., Ledger, J., Behrens, S., & Wilson, S. (2016). The impact of leadership and leadership development in higher education : a review of the literature and evidence. In *Leadership Foundation for Higher Education* (Issue April 2016).

Dornan, T. (2005). Osler, Flexner, apprenticeship and "the new medical education." *Journal of the Royal Society of Medicine*, *98*(3), 91–95. https://doi.org/10.1258/jrsm.98.3.91

Drennan, J. (2012). Masters in nursing degrees: An evaluation of management and leadership outcomes using a retrospective pre-test design. *Journal of Nursing Management*, *20*(1), 102–112. https://doi.org/10.1111/j.1365-2834.2011.01346.x

Drennan, J., & Hyde, A. (2008). Controlling response shift bias: The use of the retrospective pre-test design in the evaluation of a master's programme. *Assessment and Evaluation in Higher Education*, *33*(6), 699–709. https://doi.org/10.1080/02602930701773026

Durning, S. J., Hemmer, P., & Pangaro, L. N. (2007). The Structure of Program Evaluation: An Approach for Evaluating a Course, Clerkship, or Components of a Residency or Fellowship Training Program. *Teaching and Learning in Medicine*, *19*(3), 308–318. https://doi.org/10.1080/10401330701366796

Edmonstone, J. (2009). Evaluating clinical leadership: A case study. *Leadership in Health Services*, *22*(3), 210–224. https://doi.org/10.1108/17511870910978132

Edmonstone, J. (2013). Healthcare leadership: Learning from evaluation. *Leadership in*

Health Services, *26*(2), 148–158. https://doi.org/10.1108/17511871311319731

Embriaco, N., Papazian, L., Kentish-Barnes, N., Pochard, F., & Azoulay, E. (2007). Burnout syndrome among critical care healthcare workers. In *Current Opinion in Critical Care* (Vol. 13, Issue 5, pp. 482–488). https://doi.org/10.1097/MCC.0b013e3282efd28a

Erickson, A. S. G., & Noonan, P. M. (2018). *Utilization-Focused Evaluation In: The SAGE Encyclopedia of Educational Research, Measurement, and Evaluation*. https://doi.org/10.4135/9781506326139

Faculty of Medical Leadership and Management. (2016). *Leadership and Management Standards for Medical Professionals*. https://www.fmlm.ac.uk/sites/default/files/content/page/attachments/Leadership and Management Standards for Medical Professionals 2nd Edition - digital format_0.pdf

Faculty of Medical Leadership and Management. (2017). *FMLM Strategy: Five Year Plan 2017-2021*. https://www.fmlm.ac.uk/sites/default/files/content/page/attachments/FMLM_Strategy_2017-2021 %282%29_0.pdf

Falcone, R. E., & Satiani, B. (2008). Physician as hospital chief executive officer. *Vascular and Endovascular Surgery*, *42*(1), 88–94. https://doi.org/10.1177/1538574407309320

Fassiotto, M., Maldonado, Y., & Hopkins, J. (2018). A long-term follow-up of a physician leadership program. *Journal of Health Organization and Management*, *32*(1), 56–68. https://doi.org/10.1108/JHOM-08-2017-0208

Fenwick, S., & Hagge, E. (2007). Effective Performance Management. *Manager*, *30*(2),

85–100. https://www.leadscollaborative.ca/document/1302/Effective_Performance_Systems_Fenwick&Hagge.pdf

Fernandez, C. S. P., Noble, C. C., Jensen, E. T., & Chapin, J. (2016). Improving Leadership Skills in Physicians: A 6-Month Retrospective Study. *Journal of Leadership Studies*, *9*(4), 6–19. https://doi.org/10.1002/jls.21420

Fiedler, F. E., & Chemers, M. M. (1974). *Leadership and effective management*.

Fisher, E. S., Berwick, D. M., & Davis, K. (2009). Achieving Health Care Reform — How Physicians Can Help. *New England Journal of Medicine*, *360*(24), 2495–2497. https://doi.org/10.1056/nejmp0903923

Follett, M. P. (1949). *Freedom and co-ordination: Lectures in business organization by Mary Parker Follett*. Management publications trust.

Francis, R. (2013). *Executive summary Report of the Mid Staffordshire NHS Foundation Trust.* https://doi.org/ISBN: 9780102981476

Frank JR, Snell L, Sherbino J, E. (2015). CanMEDs 2015 Physician Competency Framework. In *CanMEDS 2015 Physician Competency Framework. Ottawa: Royal College of Physicians and Surgeons of Canada* (pp. 1–30). http://www.royalcollege.ca/portal/page/portal/rc/canmeds/resources/publications

Frich, J. C., Brewster, A. L., Cherlin, E. J., & Bradley, E. H. (2015). Leadership Development Programs for Physicians: A Systematic Review. *J Gen Intern Med*, *30*(5), 656–674. https://doi.org/10.1007/s11606-014-3141-1

Frich, J. C., & Spehar, I. (2018). Physician leadership development: towards multidisciplinary programs? *BMJ Leader*, *2*(3), 91–94. https://doi.org/10.1136/leader-2018-000070

Frugé, E., Mahoney, D. H., Poplack, D. G., & Horowitz, M. E. (2010). Leadership: “They never taught me this in medical school.” *Journal of Pediatric Hematology/Oncology*, *32*(4), 304–308. https://doi.org/10.1097/MPH.0b013e3181cf4594

Fry, L. W. (2003). Toward a theory of spiritual leadership. *Leadership Quarterly*, *14*(6), 693–727. https://doi.org/10.1016/j.leaqua.2003.09.001

Frye, A. W., & Hemmer, P. A. (2012). Program evaluation models and related theories: AMEE Guide No. 67. *Medical Teacher*, *34*(5), e288–e299. https://doi.org/10.3109/0142159X.2012.668637

Gagliano, N. J., Ferris, T., Colton, D., Dubitzky, A., Hefferman, J., & Torchiana, D. (2010). A physician leadership development program at an academic medical center. *Quality Management in Health Care*, *19*(3), 231–238. https://doi.org/10.1097/QMH.0b013e3181eb13ab

Gardner, W. L., Lowe, K. B., Moss, T. W., Mahoney, K. T., & Cogliser, C. C. (2010). Scholarly leadership of the study of leadership: A review of The Leadership Quarterly’s second decade, 2000-2009. *Leadership Quarterly*, *21*(6), 922–958. https://doi.org/10.1016/j.leaqua.2010.10.003

Geerts, J. M., Goodall, A. H., & Agius, S. (2020). Evidence-based leadership development for physicians: A systematic literature review. *Social Science and Medicine*, *246*, 112709. https://doi.org/10.1016/j.socscimed.2019.112709

Geldhof, G. J., Warner, D. A., Finders, J. K., Thogmartin, A. A., Clark, A., & Longway, K. A. (2018). Revisiting the utility of retrospective pre-post designs: The need for mixed-method pilot data. *Evaluation and Program Planning*, *70*, 83–89. https://doi.org/10.1016/j.evalprogplan.2018.05.002

General Medical Council. (2012). Leadership and management for all doctors. In *British*

Journal of General Practice (Vol. 62, Issue 598). https://doi.org/10.3399/bjgp12X636290

General Medical Council. (2018). *Outcomes for graduates 2018*. 25. https://www.gmc-uk.org/-/media/documents/dc11326-outcomes-for-graduates-2018_pdf-75040796.pdf

Gilfoyle, E., Gottesman, R., & Razack, S. (2007). Development of a leadership skills workshop in paediatric advanced resuscitation. *Medical Teacher*, *29*(9–10), e276–e283. https://doi.org/10.1080/01421590701663287

Goedhart, H., & Hoogstraten, J. (1992). The Retrospective Pretest and the Role of Pretest Information in Evaluative Studies. *Psychological Reports*, *70*(3), 699–704. https://doi.org/10.2466/pr0.1992.70.3.699

Goldstein, A. O., Calleson, D., Bearman, R., Steiner, B. D., Frasier, P. Y., & Slatt, L. (2009). Teaching Advanced Leadership Skills in Community Service (ALSCS) to medical students. *Academic Medicine*, *84*(6), 754–764. https://doi.org/10.1097/ACM.0b013e3181a40660

Goodall, A. H. (2011). Physician-leaders and hospital performance: Is there an association? *Social Science and Medicine*, *73*(4), 535–539. https://doi.org/10.1016/j.socscimed.2011.06.025

Goodall, A. H., & Stoller, J. K. (2017). The future of clinical leadership: evidence for physician leadership and the educational pathway for new leaders. *BMJ Leader*, *1*(2), 8–11. https://doi.org/10.1136/leader-2017-000010

Grint, K. (1997). *Leadership : classical, contemporary, and critical approaches*. Oxford University Press.

Grint, K. (2000). *The arts of leadership*. Oxford University Press.

Grove, J. T., Kibel, B. M., & Haas, T. (2005). EvaluLEAD: A guide for shaping and evaluating leadership development programs. In *Sustainable Leadership Initiative*.

Gulati, K., Singh, A. R., Kumar, S., Verma, V., Gupta, S. K., & Sarkar, C. (2019). Impact of a leadership development programme for physicians in India. *Leadership in Health Services*, *33*(1), 73–84. https://doi.org/10.1108/LHS-05-2019-0027

Haftel, H. M., Swan, R., Anderson, M. S., Caputo, G. L., Frohna, J. G., Li, S. T. T., Shugerman, R. P., Trimm, F., Vinci, R. J., Waggoner-Fountain, L. A., & Bostwick, S. B. (2018). Fostering the Career Development of Future Educational Leaders: The Success of the Association of Pediatric Program Directors Leadership in Educational Academic Development Program. *Journal of Pediatrics*, *194*, 5-6.e1. https://doi.org/10.1016/j.jpeds.2017.11.066

Haji, F., Morin, M.-P., & Parker, K. (2013). Rethinking programme evaluation in health professions education: beyond 'did it work?' *Medical Education*, *47*(4), 342–351. https://doi.org/10.1111/medu.12091

Hammick, M., Dornan, T., & Steinert, Y. (2010). Conducting a best evidence systematic review. Part 1: From idea to data coding. BEME Guide No. 13. *Medical Teacher*, *32*(1), 3–15. https://doi.org/10.3109/01421590903414245

Hannes, K., Lockwood, C., & Pearson, A. (2010). A comparative analysis of three online appraisal instruments' ability to assess validity in qualitative research. *Qualitative Health Research*, *20*(12), 1736–1743. https://doi.org/10.1177/1049732310378656

Hannum, K. (2004). Best practices: Choosing the right methods for evaluation. *Leadership in Action*, *23*(6), 15–20. https://doi.org/10.1002/lia.1047

Hannum, K., & Craig, S. B. (2010). Introduction to special issue on leadership development evaluation. *Leadership Quarterly*, *21*(4), 581–582.

https://doi.org/10.1016/j.leaqua.2010.06.001

Hannum, K., Martineau, J., & Reinelt, C. (Eds.). (2007). *The handbook of leadership development evaluation*. Centre for Creative Leadership.

Harden, R., Grant, J., & Buck, G. (1999). BEME Guide No. 1: Best Evidence Medical Education. *Medical Teacher*, *21*(6), 553–562. https://doi.org/10.1080/01421599978960

Harden, R. M. G., Downie, W. W., Stevenson, M., Wilson, G. M., Downie, W. W., & Wilson, G. M. (1975). Assessment of Clinical Competence using Objective Structured Examination. *British Medical Journal*, *1*(5955), 447–451. https://doi.org/10.1136/bmj.1.5955.447

Harden, R. M., Grant, J., Buckley, G., & Hart, I. R. (1999). BEME Guide No. 1: Best evidence medical education. In *Medical Teacher* (Vol. 21, Issue 6, pp. 553–562). Informa Healthcare. https://doi.org/10.1080/01421599978960

Higgins, J. (2011). Cochrane Handbook for systematic reviews of interventions version. In *The Cochrane Collaboration* (Issue September, pp. 2–3). https://doi.org/10.1002/9780470712184.ch5

Hill, L. G. (2020). Back to the future: Considerations in use and reporting of the retrospective pretest. *International Journal of Behavioral Development*, *44*(2), 184–191. https://doi.org/10.1177/0165025419870245

Hill, L. G., & Betz, D. L. (2005). Revisiting the retrospective pretest. *American Journal of Evaluation*, *26*(4), 501–517. https://doi.org/10.1177/1098214005281356

Hopkins, J., Fassiotto, M., Ku, M. C., Mammo, D., & Valantine, H. (2018). Designing a physician leadership development program based on effective models of physician education. *Health Care Management Review*, *43*(4), 293–302.

https://doi.org/10.1097/HMR.0000000000000146

Hoppe, B., & Reinelt, C. (2010). Social network analysis and the evaluation of leadership networks. *Leadership Quarterly*, *21*(4), 600–619. https://doi.org/10.1016/j.leaqua.2010.06.004

Howard, G. S., Ralph, K. M., Gulanick, N. A., Maxwell, S. E., Nance, D. W., & Gerber, S. K. (1979). Internal Invalidity in Pretest-Posttest Self-Report Evaluations and a Re-evaluation of Retrospective Pretests. *Applied Psychological Measurement*, *3*(1), 1–23. https://doi.org/10.1177/014662167900300101

Hunziker, S., Bühlmann, C., Tschan, F., Balestra, G., Legeret, C., Schumacher, C., Semmer, N. K., Hunziker, P., & Marsch, S. (2010). Brief leadership instructions improve cardiopulmonary resuscitation in a high-fidelity simulation: A randomized controlled trial. *Critical Care Medicine*, *38*(4), 1086–1091. https://doi.org/10.1097/CCM.0b013e3181cf7383

Husebø, S. E., & Akerjordet, K. (2016). Quantitative systematic review of multi-professional teamwork and leadership training to optimize patient outcomes in acute hospital settings. *Journal of Advanced Nursing*, *72*(12), 2980–3000. https://doi.org/10.1111/jan.13035

Ireri, S., Walshe, K., Benson, L., & Mwanthi, M. A. (2011). A qualitative and quantitative study of medical leadership and management: experiences, competencies, and development needs of doctor managers in the United Kingdom. *Journal of Management & Marketing in Healthcare*, *4*(1), 16–29. https://doi.org/10.1179/175330304X10Y.0000000004

Jamieson, S. (2004). Likert scales: how to (ab)use them. *Medical Education*, *38*(12), 1217–1218. https://doi.org/10.1111/j.1365-2929.2004.02012.x

Jefferies, R., Sheriff, I. H. N., Matthews, J. H., Jagger, O., Curtis, S., Lees, P., Spurgeon, P. C., Fountain, D. M., Oldman, A., Habib, A., Saied, A., Court, J., Giannoudi, M., Sayma, M., Ward, N., Cork, N., Olatokun, O., Devine, O., O'Connell, P., … Fountain, D. M. (2016). Leadership and management in UK medical school curricula. *Journal of Health, Organisation and Management*, *30*(7), 1081–1104. https://doi.org/10.1108/JHOM-03-2016-0042

Johnson, J. M., & Stern, T. A. (2014). Teaching Residents About Emotional Intelligence and Its Impact on Leadership. *Academic Psychiatry*, *38*(4), 510–513. https://doi.org/10.1007/s40596-014-0048-4

Jones, M. K., Jones, R. J., Latreille, P. L., & Sloane, P. J. (2008). Training, job satisfaction, and workplace performance in Britain: Evidence from WERS 2004. In *Institute for the Study of Labor*. https://doi.org/10.1111/j.1467-9914.2008.00434.x

Kammi, S. (1999). Stakeholder Analysis Guidelines. *Analysis*, *15*, 338–345. https://www.who.int/workforcealliance/knowledge/toolkit/33.pdf

Kane, M. T. (2001). Current concerns in validity theory. *Journal of Educational Measurement*, *38*(4), 319–342. https://doi.org/10.1111/j.1745-3984.2001.tb01130.x

Kennedy, P. E., Chyung, S. Y., Winiecki, D. J., & Brinkerhoff, R. O. (2014). Training professionals' usage and understanding of Kirkpatrick's Level 3 and Level 4 evaluations. *International Journal of Training and Development*, *18*(1), 1–21. https://doi.org/10.1111/ijtd.12023

Kirkpatrick, D. L. (1967). Evaluation of training. In R. . Craig & L. R. Bittel (Eds.), *Training and Development Handbook*. McGraw-Hill Book Company.

Kirkpatrick, D. L. (1979). Techniques for evaluating training programs. *Training and*

Development Journal, 178–192.

Klein, K. J., & Kozlowski, S. W. J. (2000). *Multilevel theory, research, and methods in organizations : foundations, extensions, and new directions*. Jossey-Bass.

Kotter, J. P. (1990). *A force for change : how leadership differs from management*. Free Press.

Krosnick, J. A., & Presser, S. (2010). Question and Questionnaire Design. In *Handbook of Survey Research* (p. 886).

Kumar, B., Swee, M. L., & Suneja, M. (2020). Leadership training programs in graduate medical education: A systematic review. *BMC Medical Education*, *20*(1), 175. https://doi.org/10.1186/s12909-020-02089-2

Kwamie, A., Dijk, van H., Agyepong, I. A., Dijk, H. van, & Agyepong, I. A. (2014). Advancing the application of systems thinking in health: Realist evaluation of the Leadership Development Programme for district manager decision-making in Ghana. *Health Research Policy and Systems*, *12*(1), 29. https://doi.org/10.1186/1478-4505-12-29

Lam, T. C. M., & Bengo, P. (2003). A Comparison of Three Retrospective Self-reporting Methods of Measuring Change in Instructional Practice. *American Journal of Evaluation*, *24*(1), 65–80. https://doi.org/10.1177/109821400302400106

Latham, G. P., & Steele, T. P. (1983). The Motivational Effects of Participation Versus Goal Setting on Performance. *Academy of Management Journal*, *26*(3), 406–417. https://doi.org/10.5465/256253

Lauer, A. K. (2018). Program & faculty evaluation. *Annals of Eye Science*, *2*(7), 44–44. https://doi.org/10.21037/aes.2017.06.02

Lempp, H., & Seale, C. (2004). The hidden curriculum in undergraduate medical education: Qualitative study of medical students' perceptions of teaching. *British Medical Journal*, *329*(7469), 770–773. https://doi.org/10.1136/bmj.329.7469.770

Leskiw, S. L., & Singh, P. (2007). Leadership development: Learning from best practices. In *Leadership and Organization Development Journal* (Vol. 28, Issue 5, pp. 444–464). https://doi.org/10.1108/01437730710761742

Leslie, L. K., Miotto, M. B., Liu, G. C., Ziemnik, S., Cabrera, A. G., Calma, S., Huang, C., & Slaw, K. (2005). Training young pediatricians as leaders for the 21st century. *Pediatrics*, *115*(3), 765–773. https://doi.org/10.1542/peds.2004-1223

Levine, S. A., Chao, S. H., Brett, B., Jackson, A. H., Burrows, A. B., Goldman, L. N., & Caruso, L. B. (2008). Chief resident immersion training in the care of older adults: An innovative interspecialty education and leadership intervention. *Journal of the American Geriatrics Society*, *56*(6), 1140–1145. https://doi.org/10.1111/j.1532-5415.2008.01710.x

Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Ioannidis, J. P. A., Clarke, M., Devereaux, P. J., Kleijnen, J., & Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *BMJ (Clinical Research Ed.)*, *339*(jul21 1), b2700–b2700. https://doi.org/10.1136/bmj.b2700

Lloyd-Smith, M. (2020). The COVID-19 pandemic: Resilient organisational response to a low-chance, high-impact event. *BMJ Leader*, *4*(3), 109–112. https://doi.org/10.1136/leader-2020-000245

Lockwood, C., Munn, Z., & Porritt, K. (2015). Qualitative research synthesis. *International Journal of Evidence-Based Healthcare*, *13*(3), 179–187. https://doi.org/10.1097/XEB.0000000000000062

Lyons, O., George, R., Galante, J., Mafi, A., Fordwoh, T., Frich, J., & Geerts, J. (2020). Evidence-based medical leadership development: a systematic review. *BMJ Leader*. https://doi.org/10.1136/leader-2020-000360

Lyons, O., McHardy, K., Bagg, W., & Wilkinson, T. (2019). Leadership development in New Zealand and Australian medical schools: Needs analysis. *BMJ Leader*, *3*(4), 123–128. https://doi.org/10.1136/leader-2019-000147

Lyons, O., Su'a, B., Locke, M., Hill, A., Su'a, B., Locke, M., & Hill, A. (2018). A systematic review of leadership training for medical students. *The New Zealand Medical Journal*, *131*(1468), 75–84. http://www.ncbi.nlm.nih.gov/pubmed/29346359

MacPhail, A., Young, C., & Ibrahim, J. E. (2015). Workplace-based clinical leadership training increases willingness to lead: Appraisal using multisource feedback of a clinical leadership program in regional Victoria, Australia. *Leadership in Health Services*, *28*(2), 100–118. https://doi.org/10.1108/LHS-01-2014-0002

Malterud, K. (2012). Systematic text condensation: A strategy for qualitative analysis. *Scandinavian Journal of Public Health*, *40*(8), 795–805. https://doi.org/10.1177/1403494812465030

Mancini, M. E., Soar, J., Bhanji, F., Billi, J. E., Dennett, J., Finn, J., Ma, M. H.-M., Perkins, G. D., Rodgers, D. L., Hazinski, M. F., Jacobs, I., & Morley, P. T. (2010). Part 12: Education, implementation, and teams: 2010 International Consensus on Cardiopulmonary Resuscitation and Emergency Cardiovascular Care Science With Treatment Recommendations [Article]. *Circulation*, *122*(16 Suppl 2), S539. https://doi.org/10.1161/CIRCULATIONAHA.110.971143

Markham, F. W., Rattner, S., Hojat, M., Louis, D. Z., Rabinowitz, C., & Gonnella, J. S. (2002). Evaluations of medical students' clinical experiences in a family medicine clerkship: Differences in patient encounters by disease severity in different clerkship

sites. *Family Medicine*, *34*(6), 451–454.

Maslach, C., & Leiter, M. P. (2016). Understanding the burnout experience: Recent research and its implications for psychiatry. *World Psychiatry*, *15*(2), 103–111. https://doi.org/10.1002/wps.20311

Maul, A. (2018). *Validity In: The SAGE Encyclopedia of Educational Research, Measurement, and Evaluation*. https://doi.org/10.4135/9781506326139

Maza, Y., Shechter, E., Pur Eizenberg, N., Segev, E. G., & Flugelman, M. Y. (2016). Physician empowerment programme; a unique workshop for physician-managers of community clinics. *BMC Medical Education*, *16*(1), 269. https://doi.org/10.1186/s12909-016-0786-y

Mckimm, J., Hickford, D., Lees, P., & Armit, K. (2019). Evaluating the impact of a national clinical leadership fellow scheme. *BMJ Leader*, *3*(2), 37–42. https://doi.org/10.1136/leader-2019-000135

McKimm, J., & Swanwick, T. (2011). Leadership development for clinicians: what are we trying to achieve? *The Clinical Teacher*, *8*(3), 181–185. https://doi.org/10.1111/j.1743-498X.2011.00473.x

McKimm, J., & Vogan, C. L. (2020). Followership: Much more than simply following the leader. *BMJ Leader*, *4*(2), 41–44. https://doi.org/10.1136/leader-2019-000162

Medical Council of New Zealand. (2014). *New Zealand Curriculum Framework for Prevocational Medical Training*. *Version 1.0*, 1–27. https://www.mcnz.org.nz/assets/News-and-Publications/NZCF26Feb2014.pdf

Meier, A. H., Boehler, M. L., McDowell, C. M., Schwind, C., Markwell, S., Roberts, N. K., & Sanfey, H. (2012). A surgical simulation curriculum for senior medical students based on TeamSTEPPS. *Archives of Surgery*, *147*(8), 761–766.

https://doi.org/10.1001/archsurg.2012.1340

Meurling, L., Hedman, L., Felländer-Tsai, L., & Wallin, C.-J. J. (2013). Leaders' and followers' individual experiences during the early phase of simulation-based team training: An exploratory study. *BMJ Quality and Safety*, *22*(6), 459–467. https://doi.org/10.1136/bmjqs-2012-000949

Militello, M., & Benham, M. K. P. (2010). "Sorting Out" collective leadership: How Q-methodology can be used to evaluate leadership development. *Leadership Quarterly*, *21*(4), 620–632. https://doi.org/10.1016/j.leaqua.2010.06.005

Ministry of Defence. (2010). *Army Doctrine: Operations*. www.dcdc.dii.r.mil.uk

Mintzberg, H. (1980). *The nature of managerial work*. Prentice Hall.

Mintzberg, H. (2004). Enough Leadership. *Harvard Business Review*, *82*(11), 22.

Monkhouse, A., Sadler, L., Boyd, A., & Kitsell, F. (2018). The Improving Global Health fellowship: A qualitative analysis of innovative leadership development for NHS healthcare professionals. *Globalization and Health*, *14*(1), 69. https://doi.org/10.1186/s12992-018-0384-3

Moreau, K. A. (2017). Has the new Kirkpatrick generation built a better hammer for our evaluation toolbox? *Medical Teacher*, *39*(9), 999–1001. https://doi.org/10.1080/0142159X.2017.1337874

Morrow, K. J., Gustavson, A. M., & Jones, J. (2016). Speaking up behaviours (safety voices) of healthcare workers: A metasynthesis of qualitative research studies. *International Journal of Nursing Studies*, *64*, 42–51. https://doi.org/10.1016/j.ijnurstu.2016.09.014

Mumford, M. D., & Van Doorn, J. R. (2001). The leadership of pragmatism:

Reconsidering Franklin in the age of charisma. *Leadership Quarterly*, *12*(3), 279–309. https://doi.org/10.1016/S1048-9843(01)00080-7

Munn, Z., Porritt, K., Lockwood, C., Aromataris, E., & Pearson, A. (2014). Establishing confidence in the output of qualitative research synthesis: The ConQual approach. *BMC Medical Research Methodology*, *14*(1), 108. https://doi.org/10.1186/1471-2288-14-108

Nakanjako, D., Namagala, E., Semeere, A., Kigozi, J., Sempa, J., Ddamulira, J. B., Katamba, A., Biraro, S., Naikoba, S., Mashalla, Y., Farquhar, C., & Sewankambo, N. (2015). Global health leadership training in resource-limited settings: A collaborative approach by academic institutions and local health care programs in Uganda. *Human Resources for Health*, *13*(1), 87. https://doi.org/10.1186/s12960-015-0087-2

Nakonezny, P. A., & Rodgers, J. L. (2005). An empirical evaluation of the retrospective pretest: Are there advantages to looking back? *Journal of Modern Applied Statistical Methods*, *4*(1), 240–250. https://doi.org/10.22237/jmasm/1114906920

National Health Service Leadership Academy. (2013). *Healthcare Leadership Model*. www.leadershipacademy.nhs.uk

National Health Service National Leadership Council. (2011). *Clinical Leadership Competency Framework*. https://www.leadershipacademy.nhs.uk/wp-content/uploads/2012/11/NHSLeadership-Leadership-Framework-Clinical-Leadership-Competency-Framework-CLCF.pdf

Neeley, S. M., Clyne, B., & Resnick-Ault, D. (2017). The state of leadership education in US medical schools: Results of a national survey. *Medical Education Online*, *22*(1). https://doi.org/10.1080/10872981.2017.1301697

NHS Improvement. (2019). Clinical Leadership - a framework for action: A guide for senior leaders on developing professional diversity at board level. *NHS Improvement*. https://improvement.nhs.uk/documents/3702/Clinical_leadership_-_framework_Jan2019.pdf

NHS Improving Quality. (2016). *NHS Improving Quality ANNEX B*.

NHS Institute for Innovation and Improvement. (2010). *Medical Leadership Competency Framework: Enhancing Engagement in Medical Leadership*. https://www.leadershipacademy.nhs.uk/wp-content/uploads/2012/11/NHSLeadership-Leadership-Framework-Medical-Leadership-Competency-Framework-3rd-ed.pdf

NHS Institute for Innovation and Improvement and Academy of Medical Royal Colleges. (2009). Medical Leadership Competency Framework. In *Leadership*.

NHS Leadership Academy. (2012). *Medical Leadership Competency Framework: Self assessment tool*. www.leadershipacademy.nhs.uk/leadership-development-module.

Nimon, K., & Allen, J. (2007). *A review of the retrospective pretest: Implications for performance improvement evaluation and research. 34*(1), 36–56.

Nimon, K., Zigarmi, D., & Allen, J. (2011). Measures of program effectiveness based on retrospective pretest data: Are all created equal? *American Journal of Evaluation*, *32*(1), 8–28. https://doi.org/10.1177/1098214010378354

Noordegraaf, M., E Schneider, M. M., J Van Rensen, E. L., & P E F Boselie, J. P. (2015). *Public Management Review Cultural Complementarity: Reshaping professional and organizational logics in developing frontline medical leadership*. https://doi.org/10.1080/14719037.2015.1066416

Norcini, J. J., & McKinley, D. W. (2007). Assessment methods in medical education. *Teaching and Teacher Education*, *23*(3), 239–250. https://doi.org/10.1016/j.tate.2006.12.021

Norman, G. (2003). Hi! How are you? Response shift, implicit theories and differing epistemologies. *Quality of Life Research*, *12*(3), 239–249. https://doi.org/10.1023/A:1023211129926

Northouse, P. G. (2013). *Leadership : theory and practice* (6th ed.). SAGE.

O'Connell, M. T., & Pascoe, J. M. (2004). Undergraduate Medical Education for the 21st Century: Leadership and Teamwork. *Family Medicine*, *36*(SUPPL.), S51–S56.

O'Neill, P. (2016). *The Leadership Development Evaluation Framework*.

Onyura, B., Crann, S., Tannenbaum, D., Whittaker, M. K., Murdoch, S., & Freeman, R. (2019). Is postgraduate leadership education a match for the wicked problems of health systems leadership? A critical systematic review. *Perspectives on Medical Education*, *8*(3), 133–142. https://doi.org/10.1007/s40037-019-0517-2

Orme, D., & Campbell, C. (2019). How leadership training saves money 'service line leadership' at Nottingham University Hospitals. *BMJ Leader*, *3*(2), 29–36. https://doi.org/10.1136/leader-2018-000132

Orvis, K. A., & Ratwani, K. L. (2010). Leader self-development: A contemporary context for leader development evaluation. *Leadership Quarterly*, *21*(4), 657–674. https://doi.org/10.1016/j.leaqua.2010.06.008

Osborn, L. M., & DeWitt, T. (2004). The HRSA-APA Faculty Development Scholars Program: Executive Leadership Track. *Ambulatory Pediatrics*, *4*(1), 98–102. https://doi.org/10.1367/1539-4409(2004)004<0098:THFDSP>2.0.CO;2

Parry, K. W., & Bryman, A. (2013). Leadership in Organizations. In *The SAGE Handbook of Organization Studies* (2nd ed.). SAGE Publ.

Patton, M. Q. (2002). *Qualitative research & evaluation methods* (3rd ed.). Sage.

Patton, M. Q. (2008). *Utilization-focused evaluation* (4th ed.). SAGE.

Patton, M. Q. (2013). *Utilization-Focused Evaluation (U-FE) Checklist*. www.wmich.edu/evalctr/checklists

Patton, M. Q. (2018). *Stakeholders In: The SAGE Encyclopedia of Educational Research, Measurement, and Evaluation*. https://doi.org/10.4135/9781506326139

Peake, L., Spurgeon, P., McKimm, J., Jones, S., Chapman, A., & Swanwick, T. (2018). *Medical leadership and management: An indicative undergraduate curriculum*.

Pearson, A., White, H., Bath-Hextall, F., Salmond, S., Apostolo, J., & Kirkpatrick, P. (2015). A mixed-methods approach to systematic reviews. *International Journal of Evidence-Based Healthcare*, *13*(3), 121–131. https://doi.org/10.1097/XEB.0000000000000052

Pearson, R. W., Ross, M. A., & Dawes, R. M. (1992). Personal recall and the limits of retrospective questions in surveys. *Questions about Questions: Inquiries into the Cognitive Bases of Surveys.*, 65–94.

Peckham, C. (2018). Medscape National Physician Burnout & Depression Report 2018. In *Medscape*. https://www.medscape.com/slideshow/2018-lifestyle-burnout-depression-6009235

Pendleton, D., & Furnham, A. (2016). *Leadership : all you need to know* (Second edi). Palgrave Macmillan.

Petrie, N. (2015). The How-To of Vertical Leadership Development – 30 Experts, 3

Conditions, and 15 Approaches. In *Center for Creative Leadership*.

Pettit, J. E., Dahdaleh, N. S., Albert, G. W., & Greenlee, J. D. (2011). Neurosurgery resident leadership development: An innovative approach. *Neurosurgery*, *68*(2), 546–550. https://doi.org/10.1227/NEU.0b013e318201c2ac

Powell, A., & Davies, H. (2016). *Managing doctors, doctors managing Research report*.

Powell, K. S., & Yalcin, S. (2010). Managerial training effectiveness: A meta-analysis 1952-2002. *Personnel Review*, *39*(2), 227–241. https://doi.org/10.1108/00483481011017435

Pradarelli, J. C., Jaffe, G. A., Lemak, C. H., Mulholland, M. W., & Dimick, J. B. (2016). A leadership development program for surgeons: First-year participant evaluation. *Surgery*, *160*(2), 255–263. https://doi.org/https://dx.doi.org/10.1016/j.surg.2016.03.011

Pratt, C. C., Mcguigan, W. M., & Katzev, A. R. (2000). Measuring program outcomes: Using retrospective pretest methodology. In *American Journal of Evaluation* (Vol. 21, Issue 3). https://doi.org/10.1177/109821400002100305

Pratt, D. D. (2002). Good Teaching: One Size Fits All? *New Directions for Adult and Continuing Education*, *2002*(93), 5–16. https://doi.org/10.1002/ace.45

Preston, C. C., & Colman, A. M. (2000). Optimal number of response categories in rating scales: Reliability, validity, discriminating power, and respondent preferences. *Acta Psychologica*, *104*(1), 1–15. https://doi.org/10.1016/S0001-6918(99)00050-5

Rabin, R. (2014). *Blended Learning for Leadership: The Centre for Creative Leadership Approach*.

Rath, T., & Conchie, B. (2008). *Strengths based leadership : great leaders, teams, and*

why people follow. Gallup Press.

Reed, D. A., Beckman, T. J., Wright, S. M., Levine, R. B., Kern, D. E., & Cook, D. A. (2008). Predictive validity evidence for medical education research study quality instrument scores: Quality of submissions to JGIM's medical education special issue. *Journal of General Internal Medicine*, *23*(7), 903–907. https://doi.org/10.1007/s11606-008-0664-3

Reed, D. A., Cook, D. A., Beckman, T. J., Levine, R. B., Kern, D. E., & Wright, S. M. (2007). Association between funding and quality of published medical education research. *Journal of the American Medical Association*, *298*(9), 1002–1009. https://doi.org/10.1001/jama.298.9.1002

Ringdahl, E. N., Tarwater, K. D., & Lindbloom, E. J. (2014). A Longitudinal Curriculum to Address the Gender Gap in Physician Leadership. *Journal of Graduate Medical Education*, *6*(2), 361–362. https://doi.org/10.4300/JGME-D-14-00081.1

Roethlisberger, F. J., Dickson, W. J., & Wright, H. A. (1939). *Management and the worker : an account of a Research Program conducted by the Western Electric Company, Hawthorne Works, Chicago*.

Rohs, F. R. (2002). Improving the Evaluation of Leadership Programs. *Journal of Leadership Education*, *1*(2), 50–61. https://doi.org/10.12806/v1/i2/rf3

Rohwer, A., Schoonees, A., & Young, T. (2014). Methods used and lessons learnt in conducting document reviews of medical and allied health curricula – a key step in curriculum evaluation. In *BMC Medical Education* (Vol. 14, Issue 1). https://doi.org/10.1186/1472-6920-14-236

Roland, D. (2015). Proposal of a linear rather than hierarchical evaluation of educational initiatives: the 7Is framework. *Journal of Educational Evaluation for Health*

Professions, *12*, 35. https://doi.org/10.3352/jeehp.2015.12.35

Rosenman, E. D., Ilgen, J. S., Shandro, J. R., Harper, A. L., & Fernandez, R. (2015). A Systematic Review of Tools Used to Assess Team Leadership in Health Care Action Teams. *Academic Medicine*, *90*(10), 1408–1422. https://doi.org/10.1097/ACM.0000000000000848

Rosenman, E. D., Shandro, J. R., Ilgen, J. S., Harper, A. L., & Fernandez, R. (2014). Leadership training in health care action teams: A systematic review. *Academic Medicine*, *89*(9), 1295–1306. https://doi.org/10.1097/ACM.0000000000000413

Rowe, M. M., & Sherlock, H. (2005). Stress and verbal abuse in nursing: Do burned out nurses eat their young? [Article]. *Journal of Nursing Management*, *13*(3), 242–248. https://doi.org/10.1111/j.1365-2834.2004.00533.x

Royal Australasian College of Medical Administrators. (2011). *RACMA Medical Leadership and Management Curriculum Framework*. www.racma.edu.au

Russon, C., & Reinelt, C. (2004). The Results of an Evaluation Scan of 55 Leadership Development Programs. *Journal of Leadership & Organizational Studies*, *10*(3), 104–107. https://doi.org/10.1177/107179190401000309

Rust, J., & Golombok, S. (2009). *Modern psychometrics: the science of psychological assessment* (3rd ed.). Routledge.

Ruston, A., & Tavabie, A. (2010). Fostering clinical engagement and medical leadership and aligning cultural values: An evaluation of a general practice specialty trainee integrated training placement in a primary care trust. *Quality in Primary Care*, *18*(4), 263–268.

Sadowski, B., Cantrell, S., Barelski, A., O'Malley, P. G., & Hartzell, J. D. (2018). Leadership Training in Graduate Medical Education: A Systematic Review. *Journal*

of Graduate Medical Education, *10*(2), 134–148. https://doi.org/10.4300/JGME-D-17-00194.1

Saks, A. M., & Burke, L. A. (2012). An investigation into the relationship between training evaluation and the transfer of training. *International Journal of Training and Development*, *16*(2), 118–127. https://doi.org/10.1111/j.1468-2419.2011.00397.x

Sarto, F., & Veronesi, G. (2016). Clinical leadership and hospital performance: assessing the evidence base. *BMC Health Services Research*, *16*, 169. https://doi.org/10.1186/s12913-016-1395-5

Satiani, B., Sena, J., Ruberg, R., & Ellison, E. C. (2014). Talent management and physician leadership training is essential for preparing tomorrow's physician leaders. *Journal of Vascular Surgery*, *59*(2), 542–546. https://doi.org/10.1016/j.jvs.2013.10.074

Savage, M., Savage, C., Brommels, M., & Mazzocato, P. (2020). Medical leadership: boon or barrier to organisational performance? A thematic synthesis of the literature. *BMJ Open*, *10*(7), e035542. https://doi.org/10.1136/bmjopen-2019-035542

Schwartz, C. E., & Rapkin, B. D. (2004). Reconsidering the psychometrics of quality of life assessment in light of response shift and appraisal. *Health and Quality of Life Outcomes*, *2*(1), 16. https://doi.org/10.1186/1477-7525-2-16

Schwartz, C. E., & Sprangers, M. A. G. (2010). Guidelines for improving the stringency of response shift research using the thentest. *Quality of Life Research*, *19*(4), 455–464. https://doi.org/10.1007/s11136-010-9585-9

Searle, J. R. (2006). Social ontology: Some basic principles. *Anthropological Theory*, *6*(1), 12–29. https://doi.org/10.1177/1463499606061731

Seidman, G., Pascal, L., & McDonough, J. (2020). What benefits do healthcare organisations receive from leadership and management development programmes? A systematic review of the evidence. *BMJ Leader*, *4*(1), 21–36. https://doi.org/10.1136/leader-2019-000141

Sharma, R., Gordon, M., Dharamsi, S., & Gibbs, T. (2015). Systematic reviews in medical education: A practical approach: AMEE Guide 94. *Medical Teacher*, *37*(2), 108–124. https://doi.org/10.3109/0142159X.2014.970996

Shaw, S., & Crisp, V. (2015). Reflections on a framework for validation – Five years on. *Research Matters; A Cambridge Assessment Publication*, *19*, 31–37. https://www.cambridgeassessment.org.uk/Images/465780-reflections-on-a-framework-for-validation-five-years-on.pdf

Sherrill, W. W. (2000). Dual-degree MD-MBA students: A look at the future of medical leadership. *Academic Medicine*, *75*(10 SUPPL.), S37-39. https://doi.org/10.1097/00001888-200010001-00012

Simon, H. A. (1947). *Administrative Behavior: A study of decision-making processes in administrative organizations*. Macmillan.

Simpson, J. G., Furnace, J., Crosby, J., Cumming, A. D., Evans, P. A., Ben David, F., Harden, R. M., Lloyd, D., McKenzie, H., McLachlan, J. C., Mcphate, G. F., Percy-Robb, I. W., & Macpherson, S. G. (2002). The Scottish doctor - Learning outcomes for the medical undergraduate in Scotland: A foundation for competent and reflective practitioners. *Medical Teacher*, *24*(2), 136–143. https://doi.org/10.1080/01421590220120713

Skeff, K. M., Stratos, G. A., & Bergen, M. R. (1992). Evaluation of a Medical Faculty Development Program: A Comparison of Traditional Pre/Post and Retrospective Pre/Post Self-Assessment Ratings. *Evaluation & the Health Professions*, *15*(3),

350–366. https://doi.org/10.1177/016327879201500307

Smith, B. W., Dalen, J., Wiggins, K., Tooley, E., Christopher, P., & Bernard, J. (2008). The brief resilience scale: Assessing the ability to bounce back. *International Journal of Behavioral Medicine*, *15*(3), 194–200. https://doi.org/10.1080/10705500802222972

Smith, K. L., Petersen, D. J., Soriano, R., Friedman, E., & Bensinger, L. D. (2007). Training tomorrow's teachers today: A national medical student teaching and leadership retreat. *Medical Teacher*, *29*(4), 328–334. https://doi.org/10.1080/01421590701316530

Solansky, S. T. (2010). The evaluation of two key leadership development program components: Leadership skills assessment and leadership mentoring. *Leadership Quarterly*, *21*(4), 675–681. https://doi.org/10.1016/j.leaqua.2010.06.009

Spehar, I., Frich, J. C., & Kjekshus, L. E. (2012). Clinicians' experiences of becoming a clinical manager: a qualitative study. *BMC Health Services Research*, *12*(1), 421. https://doi.org/10.1186/1472-6963-12-421

Spurgeon, P., Clark, J., & Ham, C. (2011). *Medical Leadership: From the Dark Side to Centre Stage*. Radcliffe Pub.

Spurgeon, P., & Down, I. (2010). *Guidance for Undergraduate Medical Education: Integrating the Medical Leadership Competency Framework Enhancing Engagement in Medical Leadership*. http://www.aomrc.org.uk/wp-content/uploads/2016/05/Undergraduate_Guidance_Medical_Education_Integrating_MLCF_0710.pdf

Spurgeon, P., Long, P., Clark, J., & Daly, F. (2015). Do we need medical leadership or medical engagement? *Leadership in Health Services*, *28*(3), 173–184.

https://doi.org/10.1108/LHS-03-2014-0029

Steinert, Y., Naismith, L., & Mann, K. (2012). Faculty development initiatives designed to promote leadership in medical education. A BEME systematic review: BEME Guide No. 19. *Medical Teacher*, *34*(6), 483–503. https://doi.org/10.3109/0142159X.2012.680937

Stogdill, R. M. (1948). Personal factors associated with leadership; a survey of the literature. *The Journal of Psychology*, *25*(1), 35–71. https://doi.org/10.1080/00223980.1948.9917362

Stogdill, R. M. (1950). Leadership, membership and organization. [Article]. *Psychological Bulletin*, *47*(1), 1–14. https://doi.org/10.1037/h0053857

Stogdill, R. M. (1974). *Handbook of leadership : a survey of theory and research*. Free Press.

Stoller, J. K. (2009). Developing Physician-Leaders: A Call to Action. *Journal of General Internal Medicine*, *24*(7), 876–878. https://doi.org/10.1007/s11606-009-1007-8

Stoller, J. K. (2020). Reflections on leadership in the time of COVID-19. In *BMJ Leader* (Vol. 4, Issue 2, pp. 77–79). BMJ Publishing Group. https://doi.org/10.1136/leader-2020-000244

Straus, S. E., Soobiah, C., & Levinson, W. (2013). The impact of leadership training programs on physicians in academic medical centers: A systematic review. *Academic Medicine*, *88*(5), 710–723. https://doi.org/10.1097/ACM.0b013e31828af493

Sturpe, D. A., & Schaivone, K. A. (2014). A primer for objective structured teaching exercises. *American Journal of Pharmaceutical Education*, *78*(5). https://doi.org/10.5688/ajpe785104

Sullivan, G. M. (2011). A Primer on the Validity of Assessment Instruments. *Journal of Graduate Medical Education*, *3*(2), 119–120. https://doi.org/10.4300/jgme-d-11-00075.1

Sultan, N., Torti, J., Haddara, W., Inayat, A., Inayat, H., & Lingard, L. (2019). Leadership Development in Postgraduate Medical Education: A Systematic Review of the Literature. *Academic Medicine*, *94*(3), 440–449. https://doi.org/10.1097/ACM.0000000000002503

Sultana, A., Sharma, R., Hossain, M. M., Bhattacharya, S., & Purohit, N. (2020). Burnout among healthcare providers during COVID-19: Challenges and evidence-based interventions. *Indian Journal of Medical Ethics*, *05*(04), 308–311. https://doi.org/10.20529/ijme.2020.73

Tasi, M. C., Keswani, A., & Bozic, K. J. (2019). Does physician leadership affect hospital quality, operational efficiency, and financial performance? *Health Care Management Review*, *44*(3), 256–262. https://doi.org/10.1097/HMR.0000000000000173

Taylor, P. J., Russ-Eft, D. F., & Taylor, H. (2009). Gilding the outcome by tarnishing the past: Inflationary biases in retrospective pretests. *American Journal of Evaluation*, *30*(1), 31–43. https://doi.org/10.1177/1098214008328517

Ten Have, E. C. M., Nap, R. E., & Tulleken, J. E. (2013). Quality improvement of interdisciplinary rounds by leadership training based on essential quality indicators of the Interdisciplinary Rounds Assessment Scale. *Intensive Care Medicine*, *39*(10), 1800–1807. https://doi.org/10.1007/s00134-013-3002-0

The King's Fund. (2011). The future of leadership and management in the NHS: No more heroes. In *London: The Kings Fund*.

Thomas, P. A., Kern, D. E., Hughes, M. T., & Chen, B. Y. (2016). *Curriculum development for medical education : a six-step approach*.

Throgmorton, C., Mitchell, T., Morley, T., & Snyder, M. (2016). Evaluating a physician leadership development program - a mixed methods approach. *Journal of Health Organization and Management*, *30*(3), 390–407. https://doi.org/https://dx.doi.org/10.1108/JHOM-11-2014-0187

Tsoh, J. Y., Kuo, A. K., Barr, J. W., Whitcanack, L., Merry, I., Alldredge, B. K., & Azzam, A. N. (2019). Developing faculty leadership from 'within': a 12-year reflection from an internal faculty leadership development program of an academic health sciences center. *Medical Education Online*, *24*(1), 1567239. https://doi.org/10.1080/10872981.2019.1567239

Uhl-Bien, M., Marion, R., & McKelvey, B. (2007). Complexity Leadership Theory: Shifting leadership from the industrial age to the knowledge era. *Leadership Quarterly*, *18*(4), 298–318. https://doi.org/10.1016/j.leaqua.2007.04.002

Uhl-Bien, M., Riggio, R. E., Lowe, K. B., & Carsten, M. K. (2014). Followership theory: A review and research agenda. In *Leadership Quarterly* (Vol. 25, Issue 1, pp. 83–104). JAI. https://doi.org/10.1016/j.leaqua.2013.11.007

van Knippenberg, D., & Sitkin, S. B. (2013). A Critical Assessment of Charismatic—Transformational Leadership Research: Back to the Drawing Board? *The Academy of Management Annals*, *7*(1), 1–60. https://doi.org/10.1080/19416520.2013.759433

Vassar, M., Wheeler, D. L., Davison, M., & Franklin, J. (2010). Program Evaluation in Medical Education: An Overview of the Utilization-focused Approach. *Journal of Educational Evaluation for Health Professions*, *7*(1), 1. https://doi.org/10.3352/jeehp.2010.7.1

Vinet, L., & Zhedanov, A. (2011). A “missing” family of classical orthogonal polynomials. *Journal of Physics A: Mathematical and Theoretical*, *44*(8), 142–148. https://doi.org/10.1088/1751-8113/44/8/085201

von Vultée, P. J., & Arnetz, B. (2004). The impact of management programs on physicians’ work environment and health: A prospective, controlled study comparing different interventions. *Journal of Health Organization and Management*, *18*(1), 25–37. https://doi.org/10.1108/14777260410532047

Vygotskiĭ, L. S., & Cole, M. (1978). *Mind in society : the development of higher psychological processes*.

Waldman, D. A., Bass, B. M., & Yammarino, F. J. (1990). Adding to Contingent-Reward Behavior: The augmenting effect of charismatic leadership. *Group & Organization Management*, *15*(4), 381–394. https://doi.org/10.1177/105960119001500404

Walumbwa, F. O., Avolio, B. J., Gardner, W. L., Wernsing, T. S., & Peterson, S. J. (2008). Authentic leadership: Development and validation of a theory-based measure. *Journal of Management*, *34*(1), 89–126. https://doi.org/10.1177/0149206307308913

Warde, C. M., Vermillion, M., & Uijtdehaage, S. (2014). A medical student leadership course led to teamwork, advocacy, and mindfulness. *Family Medicine*, *46*(6), 459–462.

Watkins, K. E., Lysø, I. H., & DeMarrais, K. (2011). Evaluating executive leadership programs: A theory of change approach. *Advances in Developing Human Resources*, *13*(2), 208–239. https://doi.org/10.1177/1523422311415643

Wayne, N. L., Vermillion, M., & Uijtdehaage, S. (2010). Gender differences in leadership amongst first-year medical students in the small-group setting. *Academic Medicine*,

85(8), 1276–1281. https://doi.org/10.1097/ACM.0b013e3181e5f2ce

Webb, A. M. B., Tsipis, N. E., McClellan, T. R., McNeil, M. J., Xu, M., Doty, J. P., Taylor, D. C., Allison M.B. Webb, MAT, Nicholas E. Tsipis, MPH, T. R. M., Michael J. McNeil, MengMeng Xu, Joseph P. Doty, PhD, and Dean C. Taylor, M., Webb, A. M. B., Tsipis, N. E., McClellan, T. R., McNeil, M. J., Xu, M., Doty, J. P., & Taylor, D. C. (2014). A First Step Toward Understanding Best Practices in Leadership Training in Undergraduate Medical Education. *Academic Medicine*, *89*(11), 1563–1570. https://doi.org/10.1097/ACM.0000000000000502

West, M., Armit, K., Loewenthal, L., Eckert, R., West, T., & Lee, A. (2015). Leadership and Leadership Development in Health Care: The Evidence Base. *The Kings Fund*, 1–36. https://doi.org/19022015

Willis, G. B. (2005). *Cognitive interviewing: a tool for improving questionnaire design*.

Wilson, A. E., & Ross, M. (2001). From chump to champ: People's appraisals of their earlier and present selves. *Journal of Personality and Social Psychology*, *80*(4), 572–584. https://doi.org/10.1037/0022-3514.80.4.572

Wilson, H. J. (2000). The myth of objectivity: is medicine moving towards a social constructivist medical paradigm? *Family Practice*, *17*(2), 203–209. https://doi.org/10.1093/fampra/17.2.203

Wormald, B. W., Schoeman, S., Somasunderam, A., & Penn, M. (2009). Assessment drives learning: An unavoidable truth? *Anatomical Sciences Education*, *2*(5), 199–204. https://doi.org/10.1002/ase.102

Wurster, A. B., Pearson, K., Sonnad, S. S., Mullen, J. L., & Kaiser, L. R. (2007). The Patient Safety Leadership Academy at the University of Pennsylvania: The First Cohort's learning experience. *Quality Management in Health Care*, *16*(2), 166–173.

https://doi.org/10.1097/01.QMH.0000267454.63123.e7

Yammarino, F. J. (1993). Transforming leadership studies: Bernard Bass' leadership and performance beyond expectations. *The Leadership Quarterly*, *4*(3–4), 379–382. https://doi.org/10.1016/1048-9843(93)90043-S

Yammarino, F. J., Spangler, W. D., & Bass, B. M. (1993). Transformational leadership and performance: A longitudinal investigation. *The Leadership Quarterly*, *4*(1), 81–102. https://doi.org/10.1016/1048-9843(93)90005-E

Yardley, S., & Dornan, T. (2012). Kirkpatrick's levels and education "evidence." *Medical Education*, *46*(1), 97–106. https://doi.org/10.1111/j.1365-2923.2011.04076.x

Young, M., & Dulewicz, V. (2008). Similarities and Differences between Leadership and Management: High-Performance Competencies in the British Royal Navy. *British Journal of Management*, *19*(1), 17–32. https://doi.org/10.1111/j.1467-8551.2007.00534.x

Yukl, G. (1989). Managerial Leadership: A Review of Theory and Research. *Journal of Management*, *15*(2), 251–289. https://doi.org/10.1177/014920638901500207

Yukl, G. A. (2013). *Leadership in organizations* (8th ed.). Pearson.

Zwack, J., & Schweitzer, J. (2013). If every fifth physician is affected by burnout, what about the other four? resilience strategies of experienced physicians. *Academic Medicine*, *88*(3), 382–389. https://doi.org/10.1097/ACM.0b013e318281696b

Appendices

Appendix One – Supplementary data table from Chapter Four

	Domain/Element satisfied (undergraduate)
	Domain/Element satisfied (foundation for postgraduate)
	Domain/Element not feasible for undergraduates
	Domain/Element not sufficiently satisfied

	MLCF Domain and description	MLCF Domain Elements	AMC Outcomes Matched
1. Demonstrating Personal Qualities	1.1 Developing self-awareness Being aware of their own values, principles and assumptions, and being able to learn from experiences	1. Recognise and articulate their own value and principles, understanding how these may differ from those of other individuals and groups 2. Identify their own strengths and limitations, the impact of their behaviour on others, and the effect of stress on their own behaviour 3. Identify their own emotions and prejudices and understand how these can affect their judgement and behaviour 4. Obtain, analyse and act on feedback from a variety of sources	4.7 **Demonstrate awareness of and explain the options available when personal values or beliefs may influence patient care,** including the obligation to refer to another practitioner. 4.9 **Self-evaluate their own professional practice;** demonstrate lifelong learning behaviours and fundamental skills in educating colleagues. Recognise the limits of their own expertise and involve other professionals as needed to contribute to patient care.
	1.2 Managing yourself Organising and managing themselves while taking account of the needs and priorities of others	1. Manage the impact of their emotions on their behaviour with consideration of the impact on others 2. Are reliable in meeting their responsibilities and commitments to consistently high standards 3. Ensure that their plans and actions are flexible, and take account of the needs and work patterns of others 4. Plan their workload and activities to fulfil work requirements and commitments, without compromising their own health	4.5 **Demonstrate awareness of factors that affect doctors' health and wellbeing, including fatigue, stress management and infection control, to mitigate health risks of professional practice. Recognise their own health needs, when to consult and follow advice of a health professional and identify risks posed to patients by their own health.**

	1.3 Continuing personal development Learning through participating in continuing professional development and from experience and feedback	1. Actively seek opportunities and challenge for personal learning and development 2. Acknowledge mistakes and treat them as learning opportunities 3. Participate in continuing professional development activities 4. Change their behaviour in the light of feedback and reflection	4.9 **Self-evaluate their own professional practice; demonstrate lifelong learning behaviours** and fundamental skills in educating colleagues. Recognise the limits of their own expertise and involve other professionals as needed to contribute to patient care.
	1.4 Acting with integrity Behaving in an open, honest and ethical manner	1. Uphold personal and professional ethics and values, taking into account the values of the organisation and respecting the culture, beliefs and abilities of individuals 2. Communicate effectively with individuals, appreciating their social, cultural, religious and ethnic backgrounds and their age, gender and abilities 3. Value, respect and promote equality and diversity 4. Take appropriate action if ethics and values are compromised	4.2 **Demonstrate professional values** including commitment to high quality clinical standards, compassion, empathy and respect for all patients. **Demonstrate the qualities of integrity, honesty,** leadership and partnership to patients, the profession and society. 4.4 **Explain the main principles of ethical practice and apply these** to learning scenarios in clinical practice. **Communicate effectively about ethical issues with patients, family and other health care professionals.**
2. Working with Others	2.1 Developing networks Working in partnership with patients, carers, service users and their representatives, and colleagues within and across systems to deliver and improve services	1. Identify opportunities where working with patients and colleagues in the clinical setting can bring added benefits 4. Actively seek the views of others 2. Create opportunities to bring individuals and groups together to achieve goals 3. Promote the sharing of information and resources	3.7 **Understand and describe the roles and relationships between health agencies and services, and explain the principles of efficient and equitable allocation of finite resources, to meet individual, community and national health needs.** **4.8 Describe and respect the roles and expertise of other health care professionals, and demonstrate ability to learn and work effectively as a member of an inter-professional team or other professional group.** **3.6 Describe a systems approach to improving the quality and safety of health care.**
	2.2 Building and maintaining relationships	1. Listen to others and recognise different perspectives 2. Empathise and take into account the needs and feelings of others	**2.1 Demonstrate by listening, sharing and responding, the ability to communicate clearly, sensitively and effectively with patients, their family/carers, doctors and other health professionals.**

	Listening, supporting others, gaining trust and showing understanding	3. Communicate effectively with individuals and groups, and act as a positive role model	**4.8 Describe and respect the roles and expertise of other health care professionals, and demonstrate ability to learn and work effectively as a member of an inter-professional team or other professional group.**
		4. Gain and maintain the trust and support of colleagues	
	2.3 Encouraging contribution Creating an environment where others have the opportunity to contribute	1. Provide encouragement, and the opportunity for people to engage in decision-making and to challenge constructively	**2.1 Demonstrate by listening, sharing and responding, the ability to communicate clearly, sensitively and effectively with patients, their family/carers, doctors and other health professionals.**
		2. Respect, value and acknowledge the roles, contributions and expertise of others	**4.8 Describe and respect the roles and expertise of other health care professionals, and demonstrate ability to learn and work effectively as a member of an inter-professional team or other professional group.**
		3. Employ strategies to manage conflict of interests and differences of opinion	**3.3 Communicate effectively in wider roles including health advocacy, teaching, assessing and appraising.**
		4. Keep the focus of contribution on delivering and improving services to patients	
	2.4 Working within teams To deliver and improve services	1. Have a clear sense of their role, responsibilities and purpose within the team	**4.8 Describe and respect the roles and expertise of other health care professionals, and demonstrate ability to learn and work effectively as a member of an inter-professional team or other professional group.**
		2. Adopt a team approach, acknowledging and appreciating efforts, contributions and compromises	
		3. Recognise the common purpose of the team and respect team decisions	
		4. Are willing to lead a team, involving the right people at the right time	
3. Managing Services	3.1 Planning Actively contributing to plans to achieve service goals	1. Support plans for clinical services that are part of the strategy for the wider healthcare system	**3.6 Describe a systems approach to improving the quality and safety of health care.**
		2. Gather feedback from patients, service users and colleagues to help develop plans	**3.8 Describe the attributes of the national systems of health care including those that pertain to the health care of Aboriginal and Torres Strait Islander peoples and/or Maori.**
		3. Contribute their expertise to planning processes	

		4. Appraise options in terms of benefits and risks	
	3.2 Managing resources	1. Accurately identify the appropriate type and level of resources required to deliver safe and effective services	3.7 Understand and describe the roles and relationships between health agencies and services, and **explain the principles of efficient and equitable allocation of finite resources**, to meet individual, community and national health needs.
	Knowing what resources are available and using their influence to ensure that resources are used efficiently and safely, and reflect the diversity of needs	2. Ensure services are delivered within allocated resources	2.5 Select and justify common investigations, with regard to the pathological basis of disease, **utility, safety and cost effectiveness**, and interpret their results.
		3. Minimise waste	
		4. Take action when resources are not being used efficiently and effectively	2.11 Prescribe medications safely, **effectively and economically** using objective evidence. Safely administer other therapeutic agents including fluid, electrolytes, blood products and selected inhalational agents.
	3.3 Managing people	1. Provide guidance and direction for others using the skills of team members effectively	**4.3 Describe the principles and practice of professionalism and leadership in health care.**
	Providing direction, reviewing performance, motivating others and promoting equality and diversity	2. Review the performance of the team members to ensure that planned services outcomes are met	3.3 Communicate effectively in wider roles including **health advocacy**, teaching, **assessing and appraising.**
		3. Support team members to develop their roles and responsibilities	4.8 Describe and respect the roles and expertise of other health care professionals, and demonstrate ability to learn and **work effectively as a member of an inter-professional team or other professional group.**
		4. Support others to provide good patient care and better services	
	3.4 Managing performance	1. Analyse information from a range of sources about performance	4.9 **Self-evaluate their own professional practice**; **demonstrate lifelong learning behaviours and fundamental skills in educating colleagues**. Recognise the limits of their own expertise and involve other professionals as needed to contribute to patient care.
	Holding themselves and others accountable for service outcomes	2. Take action to improve performance	3.3 Communicate effectively in wider roles including health advocacy, teaching, **assessing and appraising.**
		3. Take responsibility for tackling difficult issues	4.8 Describe and respect the roles and expertise of other health care professionals, and demonstrate ability to learn and **work effectively as a member of an inter-professional team or other professional group.**
		4. Build learning from experience into future plans	

4. Improving Services	4.1 Ensuring patient safety Assessing and managing the risk to patients associated with service developments, balancing economic considerations with the need for patient safety	1. Identify and quantify the risk to patients using information from a range of sources 2. Use evidence, both positive and negative, to identify options 3. Use systematic ways of assessing and minimising risk 4. Monitor the effects and outcomes of change [audit]	2.14 **Place the needs and safety of patients at the centre of the care process**. Demonstrate safety skills including infection control, graded assertiveness, adverse event reporting and effective clinical handover. **3.6 Describe a systems approach to improving the quality and safety of health care.** 2.5 Select and justify common investigations, with regard to the pathological basis of disease, utility, **safety and cost effectiveness**, and interpret their results. 2.11 Prescribe medications **safely**, effectively and **economically** using objective evidence. Safely administer other therapeutic agents including fluid, electrolytes, blood products and selected inhalational agents.
	4.2 Critically evaluating Being able to think analytically, conceptually and to identify where services can be improved, working individually or as part of a team	1. Obtain and act on patient, carer and user feedback and experiences 2. Assess and analyse processes using up-to-date improvement methodologies 3. Identify healthcare improvements and create solutions through collaborative working 4. Appraise options, and plan and take action to implement and evaluate improvements	**3.6 Describe a systems approach to improving the quality and safety of health care.** **1.4 Access, critically appraise, interpret and apply evidence from the medical and scientific literature.**
	4.3 Encouraging improvement and innovation Creating a climate of continuous service improvement	1. Question the status quo 2. Act as a positive role model for innovation 3. Encourage dialogue and debate with a wide range of people	**3.1 Accept responsibility to protect and advance the health and wellbeing of individuals, communities and populations.** **3.6 Describe a systems approach to improving the quality and safety of health care.** 1.6 Demonstrate a commitment to excellence, evidence based practice and the **generation of new scientific knowledge.**

		4. Develop creative solutions to transform services and care	4.9 **Self-evaluate** their own professional practice; **demonstrate lifelong learning behaviours** and fundamental skills in **educating colleague**s. **Recognise the limits of their own expertise** and involve other professionals as needed to contribute to patient care.
	4.4 Facilitating transformation Actively contributing to change processes that lead to improving healthcare	1. Model the change expected	**1.6 Demonstrate a commitment to excellence, evidence based practice and the generation of new scientific knowledge.**
		2. Articulate the need for change and its impact on people and services	**3.6 Describe a systems approach to improving the quality and safety of health care.**
		3. Promote changes leading to systems redesign	**3.3 Communicate effectively in wider roles including health advocacy, teaching, assessing and appraising.**
		4. Motivate and focus a group to accomplish change	
5. Setting Direction	5.1 Identifying the contexts for change Being aware of the range of factors to be taken into account	1. Demonstrate awareness of the political, social, technical, economic, organisational and professional environment	**3.2 Explain factors that contribute to the health, illness, disease and success of treatment of populations, including issues relating to health inequities and inequalities, diversity of cultural, spiritual and community values, and socio-economic and physical environment factors.**
		2. Understand and interpret relevant legislation and accountability frameworks	**3.9 Demonstrate an understanding of global health issues and determinants of health and disease including their relevance to health care delivery in Australia and New Zealand and the broader Western Pacific region.**
		3. Anticipate and prepare for the future by scanning for ideas, best practice and emerging trends that will have an impact on health outcomes	**3.4 Understand and describe the factors that contribute to the health and wellbeing of Aboriginal and Torres Strait Islander peoples and/or Māori, including history, spirituality and relationship to land, diversity of cultures and communities, epidemiology, social and political determinants of health and health experiences.** Demonstrate effective and culturally competent communication and care for Aboriginal and Torres Strait Islander peoples and/or Māori.
		4. Develop and communicate aspirations	
	5.2 Applying knowledge and evidence	1. Use appropriate methods to gather data and information [audit]	**1.4 Access, critically appraise, interpret and apply evidence from the medical and scientific literature.**

	Gathering information to produce an evidence-based challenge to systems and processes in order to identify opportunities for service improvements	2. Carry out analysis against an evidence-based criteria set	**1.5 Apply knowledge of common scientific methods to formulate relevant research questions and select applicable study designs.** **1.6 Demonstrate a commitment to excellence, evidence based practice and the generation of new scientific knowledge.**
		3. Use information to challenge existing practices and processes	3.5 **Explain and evaluate common population health screening and prevention approaches, including the use of technology for surveillance and monitoring of the health status of populations.** Explain environmental and lifestyle health risks and advocate for healthy lifestyle choices.
		4. Influence others to use knowledge and evidence to achieve best practice	**3.6 Describe a systems approach to improving the quality and safety of health care.**
	5.3 Making decisions Using their values, and the evidence, to make good decisions	1. Participate in and contribute to organisational decision-making processes	4.7 Demonstrate awareness of and explain the options available when **personal values or beliefs may influence patient care**, including the obligation to refer to another practitioner.
		2. Act in a manner consistent with the values and priorities of their organisation and profession	2.7 **Make clinical judgements and decisions based on the available evidence.** Identify and justify relevant management options alone or in conjunction with colleagues, according to level of training and experience.
		3. Educate and inform key people who influence and make decisions	4.4 **Explain the main principles of ethical practice and apply these** to learning scenarios in clinical practice. Communicate effectively about ethical issues with patients, family and other health care professionals.
		4. Contribute a clinical perspective to team, department, system and organisational decisions	4.2 **Demonstrate professional values** including commitment to high quality clinical standards, compassion, empathy and respect for all patients. Demonstrate the qualities of integrity, honesty, leadership and partnership to patients, the profession and society.
	5.4 Evaluating impact	1. Test and evaluate new service options [audit]	**4.9 Self-evaluate their own professional practice; demonstrate lifelong learning behaviours** and fundamental skills in educating colleagues. Recognise the limits of their own expertise and involve other professionals as needed to contribute to patient care.

	Measuring and evaluating outcomes, taking corrective action where necessary and by being held account for their decisions	2. Standardise and promote new approaches	3.5 Explain and **evaluate common population health screening and prevention approaches**, including the use of technology for surveillance and monitoring of the health status of populations. Explain environmental and lifestyle health risks and advocate for healthy lifestyle choices.
		3. Overcome barriers to implementation [important to be aware of]	1.5 Apply knowledge of common scientific methods to **formulate relevant research questions** and select applicable study designs.
		4. Formally and informally disseminate good practice	**3.6 Describe a systems approach to improving the quality and safety of health care.**

Table 16: Summary of matching between the Medical Leadership Competency Framework and the Australian Medical Council Outcomes for Graduates.

Appendix Two – Supplementary data for Chapter Ten

Also available at:
https://bmjleader.bmj.com/content/leader/suppl/2020/11/16/leader-2020-000360.DC1/leader-2020-000360supp001_data_supplement.pdf

Source (First Author, Year)	Setting	Learner Number	Learner Type	Intervention Length	Intervention Description	Teaching Methods	Educational Content	Main findings by Kirkpatrick level	JBI Score	MERSQI Score
Higher Reliability Studies (MERSQI)										
Boyle, 2004 [1]	Two US ICUs	10	3 Physicians 7 Nurses	8 months	6 externally provided modules, total 23.5hrs	Learning activities, small group skill practice, problem-solving sessions, feedback and reinforcement of skills, assignment, assessments, feedback	Leadership, communication, coordination, problem solving/conflict management, and team culture	3a. Leaders reported increased satisfaction with their own communication and leadership skills 3b. Communication skills of ICU nurse and physician leaders improved significantly in simulation (from 57 to 75/100). Relationship skills remained high (77-78/100). 4b. Reported increased problem-solving between groups and decreased personal stress in one of the sites.	n/a	16
Parsons, 2018 [2]	Single US hospital residency	14	Residents (Emergency Medicine)	4 days	4 days of simulation scenarios	Introductory didactic presentation followed by a series of 6 simulation scenarios and structured debriefs.	Crew Resource Management (CRM)	2b. Each team showed an overall gradual improvement in CRM skills compared to the preceding teams, suggesting that observational learning of CRM was effective in this setting. 3b. Large but not significant increases in all objective measures of leadership, problem solving, situational awareness, resource utilisation, communication, and overall crisis resource management score (overall 2.75/7 to overall 6.0/7). Very low number of teams (4) caused lack of significance (4)	n/a	14.5
Cooper, 2001 [3]	UK advanced life support course	35	Mixed seniority doctors, nurses and technicians	3 day resuscitation course	75-min leadership development seminar	Lectures, videos and discussion groups, home reading	Importance of leadership, behaviours of effective leadership, introversion/extroversion	1. Appraisal of the leadership seminar was very positive (mean 4/5) 2a. Reported increased confidence in role as a leader 3b. Significant improvement over the control group 9/10 items on the leadership observation. Mean increase of 4.53(/40) (cf. 2.23 in the control group)	6	14
Malling, 2009 [4]	Single educational region in Denmark	28	Consultants (Responsible for education)	6 months	Two three-day residential modules and a follow-up day.	Residential modules and follow-up day. Mandatory assignments.	Pedagogical knowledge, organization of specialist training, educational culture evaluation and quality assurance, planning specialist training in the department, supervision of supervisors, implementation strategies, personal development, leadership in specialist training, research in medical education	1. Participants rated the course as beneficial and meeting their expectations (3.2-3.3/4) 2b. Technical, administrative and human skills feedback did not improve or differ from the control group. 3b. Citizenship behaviours did not improve or differ from the control group	n/a	14
von Vultée, 2004 [5]	University hospitals in Sweden	52	Specialists, senior physicians, heads of departments	1 year	Three programs, including mentor programs, management networks, and lectures held across 1 year; no details on number, duration, or allocation to programs	Mentoring/Networking/Lectures	NR	2a. No significant differences in self-reported well-being, self- esteem, mental energy, influence, authority, efficiency, assessed using elements of the quality, work, competence tool (data not provided) 2b. No significant differences in self-reported skills development, self-esteem, mental energy or work-related exhaustion 3a. No significant differences in influence, authority, participation, feedback, goal clarity or efficiency 3b. No significant difference in senior management positions between program and control 4b. Sick leave increased by 6.9 days per year fewer in intervention group compared to reference group (1.3 days vs 8.2 days), p<0.05	n/a	14
Fassiotto, 2018 [6]	Single US hospital	131	Assistant/associate/full professors	9 months	6 x 1.5 day sessions over 9 months	Interactive teaching methods based on adult earning principles, action learning projects	Personal development as a leader managing people and relationships managing groups and projects managerial finance and accounting understanding the organizational system	1. Positive qualitative feedback about the course 2a. Increased perceived institutional support (no Bonferroni) 2b. Self-reported increased understanding of organisational structure esp. Finance 3b. Participants more likely to hold regional or national leadership titles and to have taken on new leadership titles. No significant difference in promotions 4b. Increased retention of female participants	3	13.5
Levine, 2008 [7]	Single US academic medical centre	47	Residents (Chief residents, medicine and surgery)	1 year	Two-day offsite immersion training, project work	Small-group discussions, evidence-based mini-lectures, interactive seminars, one-on-one project mentoring	Management of complex older patients, geriatric principles, giving feedback, approaching the reluctant learner, conflict resolution	1. Effectiveness of programme rated at 3.69/5 2b. Reported increased confidence in skills and knowledge in role of chief resident. Knowledge test significantly increased in 2 of 3 years. Significantly increased self-assessed knowledge 3a. Reports of heightened sensitivity to the unique needs of older patients 3b. Eight individuals accomplished 100% of their projects, 20/27 completed at least half of the project.	0	13.5
Hopkins, 2018 [8]	Single US hospital network	113	Senior medical leaders and academic faculty. 19 administrators, 94 doctors.	9 months	Six 1.5 day sessions spaced over 9 months	Baseline assessments of their leadership competence, Multi-Source Feedback, Myers Briggs Type Indicator and the Thomas-Kilmann Conflict Mode Instrument. Reading materials, assignments, case study, role-play, discussions in dyads, brief reflection and writing assignments, responses to video vignettes, brainstorming, and small group problem-solving assignments, with minimal	Personal development as a leader Managing people and relationships Managing groups and projects Managerial finance and accounting Understanding the organizational system	1. Programme rated 4.5/5 overall 2a. Significant improvements in self-reported attitudes 2b. Significant improvements in self-reported knowledge and skills 3a. Self-reported significant improvement in effectiveness as a leaders and power and influence 3b. 100% of participants completed their projects. 4b. 58% of projects achieved IHI level 3 (moderate improvement in process measures) with 22% of these attaining level 4 (significant improvement in outcomes measures).	n/a	13.5

					emphasis on didactic lectures. Project work					
Dannels, 2008 [9]	Single US university executive education programme	78	Female academic medical faculty (associate or full professor level)	1 year	Executive leadership development program for senior female faculty	Executive leadership education	Not specified	2a. Aspiration to higher leadership position inside an academic health centre decreased; 2b. In all eight leadership composites, the exposure group mean (based on a seven-point scale) was slightly but significantly greater than the means in both the control groups. (average 0.2/7 increase) 3b. A higher percentage of participants have achieved leadership positions 4b. Promotion of increased number of female graduates achieved	n/a	13
Orme, 2019 [10]	Single UK hospital trust	425	Consultants and senior healthcare professionals and managers	12 months	12 months with 5 workshop days spread over 6 months	Five face-to-face delivery days, ongoing tele-phone coaching and the use of a benchmarked 360-degree profile. Project work. Interactive sessions. 3 days, then 1 day after 3 months, then 1 day after 6 months. Support via coaching between. 360 repeats at 12 months	Not well reported: influence, behaviours, team behaviours, finances. From website: execution, people management, evaluation and feedback	1. Evaluation using the net promotor score gave 92% score (promotors-detractors/total) 2a. Reports of improved confidence 2b. Reports of improved skills and techniques, self-awareness and other-awareness. 3a. Reports of improvement in delegation skills, trust. Corroboration with interview questionnaire 3b. Multi source feedback showed statistically significant improvement in clarifying purpose, with several other categories approaching significance. 4a. Reports of improved team effectiveness. 4b. Savings of £3.3 million were identified through the delivery of 11 separate initiatives for a Return On Investment of 364%. Quality improvements were observed.	6	12.5
Cole, 2017 [11]	Single US hospital department	10	Residents (Anaesthesiology PGY3/4)	2 weeks	2 week operating room management and leadership elective rotation	Action learning literature reflective learning	Non-technical skills	3a. Increased "re-evaluates and debriefs"; "gathers and actively seeks out information"; "anticipates changing environment" 3b. Increase across a range of metrics measuring teamwork, task management, clinical decision making, situational awareness, as measured by "anaesthetists' non-technical skills" (ants) questionnaire	n/a	12.5
Haftel HM, 2018 [12]	Single US specialty association, 45 sites	49	Paediatric academic faculty	10 months	3 sessions focusing on the individual, their training programme and interaction with others	"highly interactive format", peer mentorship	Professional development, leadership training, administrative skill development.	3b. Statistically significant increase in leadership of national committees, production of national workshops and presentation at national platforms.	n/a	12.5
Ten Have, 2013 [13]	Four Dutch ICUs (in Single hospital)	9	Exposure: intensive care fellows Control: experienced intensivists	23 months	1 day simulation, group feedback on videoed interdisciplinary rounds. (unclear when this was offered - before or after post-training video)	Multiple learning activities including simulation; small group skill practice and problem-solving sessions; performance feedback and reinforcement of newly learned skills; and a planning assignment for on the job applications.	Leading an interdisciplinary ward round	3b. Participants increased significantly in their performance of 7 of the 10 tasks on the interdisciplinary ward round leadership assessment tool. Post-test, the participants performed these behaviours significantly more frequently than an experienced control group.	n/a	12.5
Gilfoyle, 2007 [14]	Single Canadian residency program (paediatrics)	29	Residents (Paediatrics, PGY1–PGY4)	1/2 day	Half-day workshop	Plenary session followed by two simulated resuscitation scenarios	Tasks required of a leader, effective communication skills within a team, and avoidance of fixation errors	2b. Significantly increased knowledge of tasks and fixation errors, "greater understanding of the concepts of effective leadership and team functioning" 3b. Residents' performance significantly improved from scenario 1 to scenario 2 (63% vs. 82%, p< 0.05). Residents' scores were better during the first scenario of the initial workshop than those during the 6-month workshop who had never previously participated (control). (63% vs. 50%, p< 0.05).	n/a	12
LoPresti, 2009 [15]	Four US residency programs	6	Residents (Family Medicine, PGY2)	2 years	60 hours of education in 20 modules.	Lectures, project work, in class exercises.	Leadership, quality improvement, policies, strategy and markets, insurance, finances, professional success, negotiation	1. Mean attendance of 66% 2b. Significant but small increase in test scores compared with control group	n/a	12
Wurster, 2007 [16]	Single US department of surgery	42	Surgical fellows	6 months	Long weekend of didactic study; teamwork on patient safety-related project; monthly conferences; 2 days for lectures and project presentations	Didactic study, group projects with monthly conferences and project presentations and capstone lectures	Ability to understand cognitive processes and group dynamics underlying medical decision making; communication across patient care continuum; implementation of systems approach to patient care	1. Reported leadership academy programme was more valuable than other patient safety initiatives 2a. Improved attitudes towards leadership roles 2b. Increased perceived leadership capability and knowledge across multiple skills and abilities 3a. Increased perceived functional skills in 5/8 areas. Increased preparedness to take a leadership role 4b. 1 out of 6 projects fully implemented	n/a	12
Higher Reliability Studies (JBI tool)										
Pradarelli, 2016 [17]	Single US hospital department	21	Academic surgeons from assistant to full professor grade	8 months	1 full day per month	Didactic and experiential learning. Case studies, team improvement projects, multi-source feedback, debriefing with executive coach	Leadership, team building, business acumen, and health care context	1. Participants reported high levels of satisfaction with the programme, ranked a 8.7/10 (10 being excellent use of their time) 2a. Participants felt "not only enabled but also capable of effecting change in their local environments" 2b. Participants reported increased self-awareness and increased team-building skills, and improvement of leadership knowledge 3a. Participants reported improved ability to foster collaborative relationships, and general improvement of interactions and networks.	9	8
Throgmorton, 2016 [18]	US Regional healthcare system	21	Physicians across a range of specialties	10 months	2.5h meetings/month +/- 2-3h of additional learning opportunities	Behavioural style assessment, multisource feedback, coaching, online discussions, online learning resources, team project in small groups	Intra/interpersonal effectiveness; resiliency; coaching; communication; teamwork; change management; business acumen; quality focus	1. Ratings of 4/5 for evaluation of eight content sessions (lowest average 4.3) 2a, 2b. Completing disc assessment, 360 feedback and coach supported development plan 3a. 16/21 participants completed individual development plan; 21/21 completed the everything disc workplace profile 3b. Team presentations completed 4B: 106% return-on-investment calculated from "intangible benefits"	8	11.5
Bergman, 2009 [19]	Single Swedish hospital	53	Managers (9 physicians, 33 nurses, and 11 other health personnel)	7 week vs 17 months	1: One week intensive course 2: Long-term support group (previously	A one-week course and a long-term support group. The intensive course consisted of modules using reflection and metre reflection	Group dynamics, communication, leadership theories	1. Participants emphasise the importance of the group as a "protected zone" 2a. Participants felt that they dared to be clearer. Both groups had improved attitude to leadership roles 3a. Participants from all groups reported using techniques in their workplace and	8	10

					completed intensive) 3. Long-term support group (had not completed intensive)	supported by theoretical frameworks. The long term support group met half a day once a month for 1 to 2 years, to discuss problems that arise in the everyday work.		personal lives, participants in the support groups reported exploring ways to handle changes in their work		
Monkhouse, 2018 [20]	National UK programme	111	Doctors (secondary and primary care), nurses, public health professionals, allied health professionals, managers	3-9 months	3-9 month placement in a resource-poor country	Formal training, identification of learning needs, mentoring, project work with overseas partner	NHS healthcare leadership model, not otherwise specified	1. All interviewees agreed that the experience was valuable 2a. Increased percentage considering themselves to be leaders after the programme (78% after, 32% before), interviewees reported increased confidence. 2b. Increased self-awareness and leadership skills reported in questionnaire. Interviewees reported increased awareness of the leadership styles of others. 3a. 73% reported using their new skills. 18% reported not being able to use their skills. Interviewees reported changes in the way they relate to others. 4a. Several respondents noted they had planned to leave the NHS and decided to stay after the programme.	8	8.5
Tsoh JY, 2019 [21]	Single US academic healthcare centre	136	Faculty members perceived to have leadership potential	20 weeks	10 leadership modules over 20 weeks (75 Programme Hours)	Experiential learning NOS	Self-awareness, critical thinking, effective communication, inclusion, collaboration, empowered professionalism	1. Programme completion rate is 97.8% 2a. 76.4% of respondents said they developed an improved attitude towards their role and/or job at the university. Qualitative comments indicated improved confidence and aspiration. 2b. 98.6% of respondents reported noticeable changes in leadership skills 3a. 91.7% of respondents reported changes in leadership attitudes or behaviours. 62.5% reported seeking new leadership opportunities. 3b. 9.6% of graduates reported a new leadership position; 12.4% of female graduates, 33.3% of underrepresented minority graduates.	8	8.5
Bearman, 2012 [22]	Single Australian residency program (surgery)	12	Residents (Surgical trainees, midlevel across all specialties)	2 days	Two day simulation course	Simulation, peer observations, multi-source feedback, reflection, lectures, videos, scenarios	Patient-centred communication, inter-professional communication, teamwork, leadership and professionalism	1. All participants rated the course as good or very good. One third of participants described the communication scenario as "less than useful". All other aspects of the course were considered useful or highly useful. 2a. Increased awareness of the broader situation and the value of high-quality communication and teamwork 2b. Self-reported achievement of learning objectives including increased knowledge	8	7.5
Carney, 2015 [23]	12 US Primary Care residencies (4 locations)	33	Faculty (Family medicine, internal medicine, Paeds)	6 months	2.5day session with followup over 6 months	Didactic small-group sessions webinars conference calls visits by core faculty	Leadership change management teamwork population management clinical microsystems competency assessment patient-centredness and patient-centred-medical-home principles	1. Participants satisfied 51.5%-97%, useful 42.5%-84.8% on 7 topics 2a. 50%-96.7% report intention to implement	8	7.5
Cooper, 2011 [24]	Single US academic medical centre	108	Physicians, nurses, allied health professionals, administrators, managers	1 day	Workshop	Seminar simulation review of data from safety climate survey team project	Teamwork, patient safety, communication, individual and collective leadership, recognizing difference in perspectives between managers and clinicians, how to speak up to voice concerns, specifically	1. Scores for relevance and quality of simulations on questionnaire and free-text comments rated >5/10. 2a. The simulation helped participants recognize problems with speaking up. 2b. Gained understanding about shifting from blame to learning oriented leadership, facilitating communication and teamwork, being welcoming rather than defensive, and other self-reflections.	8	6.5
Agius, 2015 [25]	Single UK deanery	8	Specialty trainees (psych, renal, GUM, ENT)	4 years	Four-year part-time programme to Master's level with academic and vocational components	Diploma/MSc modules Action learning sets Workplace-based projects Shadowing placements Leadership development tools (self-reflection)	MLCF competencies Leadership, policy, organisational development, governance	1. Participants were happy with the course 2a. Participants reported improved confidence and judgement skills 2b. Participants reported improved leadership skills 3a. Participants reported taking new approaches to their roles 3b. Participants reported new roles 4b. Projects had significant system impact including organisational changes, improved patient experiences, cost savings.	7	10.5
McKimm J, 2019 [26]	National UK programme	145	Residents (Specialty and GP doctors in training)	1 year	Immersive internship out of practice with the most senior personnel in national and healthcare-related organisations	Immersive internship, visits to other host organisations and Parliament, teaching on leadership and management, and action learning sets	Policy development, project management, research and analysis, writing and publishing, professional networking skills.	1. Participants overwhelmingly endorse the programme and would recommend it to other trainees. 2a. Increased self-confidence as leaders (87% of respondents) and willingness to speak up 2b. Increased understanding of health systems, policy, team working skills with diverse groups, networking 3a. 60% of graduates report being actively involved in service improvements post-programme 48% of graduates reported difficulties transferring their training back to their clinical practice. 3b. 63% of graduates had gone on to further leadership/management development as a result of the programme 4b. Host organisations describe a range of benefits and examples of the impact of fellows' work on their organisations, including financial impact (for example, Income generation, cost savings) and a range of deliverables (for example, Reports, publications, research studies).	7	10.5
Cohen, 2017 [27]	Multiple UK health care organizations (London)	69	Consultants (32), Registrars (35), Grade not identified (2)	1 day	Workshop	Lecture-style presentations policy leadership simulation	Background of NHS reform healthcare challenges	1. All measures above 4/5 on Likert scales including "I recommend this simulation to colleagues at my professional stage" (4.50) and "I would like to take part in similar simulation events in future". 2a. Self-reported "learning from the simulation will help me deliver better long term care to my patients" (4/5) 2b. Self-reported increase in understanding of healthcare organisation (4.55/5) in all seven knowledge areas (regulation of health care providers, role of patient organizations, organizational accountability, role of local authorities, financial climate, roles and responsibilities of commissioners and care providers), there was a significant improvement in perceived knowledge scores. 3a. Capability was the only construct that showed a large and significant change post-	7	10

								simulation. There was a moderate but significant change in behavioural intentions, attitudes and subjective norms post simulation, although there was no significant change detectable in opportunity in the short time between the simulation and feedback, 22 clinicians stated that their experiences in the crucible simulation had directly influenced their leadership practice increase in perceived self-efficacy post simulation (mean score pre simulation, 3.87 to post simulation, 4.08).		
Ruston, 2010 [28]	Single UK deanery	3	Residents (General Practice Specialty Trainees)	4 months	2 days per week for 4 months	Peer learning sets, meetings, project work, reflective diaries	Strategic and contextual issues, commissioning, design and delivery of health care, service redesign, public health agenda, leadership, management and partnership skills.	1. All respondents were positive about the value and success of the programme 2b. "evidence indicated that the [trainees] had met their core curriculum competencies at the level expected.". Trainees reported gaining an understanding of how good leaders hold a vision of the future and motivate people. All respondents, including observers, reported a marked increase in the trainee knowledge. Increased self-awareness 3a. Trainees reported using the skills learned during their project work 3b. Observers reported change in trainee behaviour, use of language, and use of skills. Supervisors also reported cultural barriers and organisational silos being broken down.	6	10.5
Other Included Studies										
Hackworth J, 2018 [29]	Single US hospital	99	Medical faculty members with supervisory or team leader responsibility	10 months	68 hours of learning over 10 month course. Two one-day seminars, one two-day off-site session, seven monthly four hour meetings, reading assignments.	Seminars, experiential activities, small group discussions, multisource feedback, teambuilding activities, reading	Emotional intelligence, leadership behaviours, leadership foundation skills	1. 94% of participants would recommend the programme to others, training rated as worthwhile investment (6.7/7) 2a. Statistically significant increased confidence in leadership ability 2b. Statistically significant improvement in 25/26 leadership measure 3a. Statistically significant increase in seeking ways to maximise application of leadership strengths and to overcome leadership limitations 3b. Statistically significant improvement in communication of leadership strengths and in confidence as seen by managers	3	11.5
Al-Mutawa, 2016 [30]	Family medicine residency programme Qatar	39	Residents (family medicine year 1-4)	5 days	5 day practice management workshop	Didactic teaching case studies small group discussions team exercises	Professionalism interpersonal skills practice-based learning and improvement system-based practice	2a. Increased confidence 2b. Increased self-assessed risk management, conflict management, communication skills, time management, ability to write objectives 3b. Supervisor-reported increase in effective use of hospital resources, coordination of patient care, patient communication skills	n/a	11.5
Fernandez, 2016 [31]	Single US college (O&G)	37	Obstetricians (Junior fellows, young physicians, senior fellows, not otherwise specified)	3.5 days	3.5 day national intensive leadership development for O&G physician leaders	Interactive skills-building workshops, series of leadership and psychological assessment tools, including a 360-degree assessment. Formal coach debrief of assessment, small and large group sessions	Organisational culture, leading and empowering, communication, motivation, advocacy, media, negotiation skills, health policy	1. 100% of respondents indicated that they would recommend the course to colleagues 2b. Post test scores were significantly and meaningfully higher than pre-test at the p< .0001 level in all 10 targeted leadership skills, both immediately following completion of the course and at 6 months post completion. Mean differences ranged from 0.8 to 1.81 (5 point scale from unskilled to highly skilled) 3a. Respondents reported having used skills learned in their day to day job. 3b. 9/26 respondents had expanded leadership responsibilities in a new role.	4	11
Chang, 2019 [32]	Three US healthcare organisations in 24 US states	65	49 Physicians (46 Geriatrics) 1 Dentist 3 Nurses 6 Pharmacists 1 Psychologist 1 Social Worker 1 Speech Pathologist	9 months	55hrs	Orientation, large-group workshop, two in-person meetings, monthly small group videoconferences disc personality type instrument, 1:1 coaching, structured networking, independent reading and reflection individual leadership project	Adaptive leadership managing reactivity courage and generosity mission, vision, goals disc(r)assessment motivation and resilience managing transitions influence and persuasion intrinsic and extrinsic rewards stories and advice presentation skills messaging and marketing strategy and stakeholders evaluation design scaling up business case and budget measuring impact project management	1. Mean satisfaction score of 4.86/5 2a. Significant increases in confidence in all sub-domains: self-awareness self-management empathy communication planning and execution (overall mean increased from 5.8/9 to 8.0/9) 2b. Some scholars stated that they gained confidence and skills from the program. 3a. Scholars reported intentionally continuing as a sustaining community of national colleagues after the conclusion of the program, offering peer mentoring, sharing speaking opportunities, and co-authoring manuscripts 3b. Scholars gave 85 presentations and published 63 manuscripts, and reported accepting local and national leadership positions as a result of the programme. 4b. Scholars received 21 awards, in addition to receiving funding and program cost savings from their training and practicum.	1	11
Day, 2010 [33]	Single US specialty association (orthopaedics)	100	Orthopaedic surgeons	1 year	Mentoring by established orthopaedic leader	Mentoring	Unclear	2b. Significant increase in 3/8 leadership domains: knowledge of theory, tolerance for demands of leadership, and leadership positioning. 3b. Significant increase in the number of national committee chairs (22% pre, 62% post-programme, p < 0.001). 48% of alumni advanced in academic rank vs 21% of controls (p = 0.005).	n/a	11
Korschun, 2007 [34]	Single US academic medical centre	70	Physicians (29), nurses, and a wide range of administrators	5 months	Five three-day sessions over five months	Lectures, seminars, case studies, experiential exercises, individual assessment, executive coaching, including a 360° assessment, mentoring, team project work	Strategic thinking and personal awareness, Leadership qualities, Leadership best practices, negotiating tactics and managing conflict Human resources and talent management, Building collaboration and influence skills, Marketing, development and public policy, Leadership in changing times, Communications, media relations, crisis management	1. Participants reported positive experiences with the program. Attendance was at a rate of 95% or higher. 24% of respondents disagreed that the mentoring process had helped them with their professional growth. 2a. 98% increased their commitment to and support of the vision and strategies of the organisation 2b. All respondents agreed to some degree that they had improved their interpersonal skills related to team leadership and that they had improved their approach to functioning in a team setting. Participants reported increased knowledge of the organisation. 3a. 93% reported the programme has made them more effective leaders. 85% report that they have become advocates for the organisation's strategy. Graduates reported being more effective in committees within their school or the university. 3b. 15% of graduates were promoted within the follow- up period. 76% reported taking on additional leadership responsibilities. 4a. Fellows stated that the leadership academy allowed them to establish a strong network of friends and colleagues from other disciplines, and after the program was over, they found it much easier to seek advice or establish collaborations with peer leaders in other parts of the organisation. 96% of respondents reported being more likely to stay with the organisation. 4b. Successful adoption of most projects into the organisation.	n/a	11
McDade, 2004 [35]	Multiple US and Canada Academic Medical Centres (AMCs)	79	Associate or full professor	7 months	Three 1-week meetings across 7 months	Leadership skill development, mentoring, and networking	The curriculum focuses on building knowledge and skills in seven domains: paradigms of corporate, government, and academic leadership; financial management; strategic planning and organizational transformation;	2a. Increase confidence in applying conflict resolution strategies, and with public speaking. 2b. Significant increases (p<0.001) in knowledge of organization, management, and leadership theory, financial management, environmental scanning, job negotiations, risk management, life balancing	n/a	11

							emerging issues in academic medicine; communication; personal dimensions of leadership; and career advancement strategies.	3a. Increase in reports of having a mentor 3b. No clear effect on career progression. 4b no significant increase in tangible benefits from projects.		
Edmonstone, 2011 [36]	UK National programme (Scotland)	117 (from total 5 cohorts)	Senior strategic leaders in NHS Scotland: hospital doctors, primary care doctors, nurses, AHPs, pharmacists, psychologists, dentists, paramedics	12 months	A three-day residential development centre followed by nine two-day residential events held every six weeks.	Coaching, Mentoring, Action Learning, Change Management Project, Masterclasses/Workshops, After Dinner Speakers, Shadowing, Organisational Visits, Chief Executive Sponsorship, multi-source feedback, psychometric tools, Development Centre, Personal Development Plan	Self-leadership, leading others, collaborative working, delivering excellence through others, managing complex change, improving patient experience, political awareness, strategic dexterity, aligning agendas/creating culture	1- all elements of the programme were positively rated by participants with ratings becoming more positive as the programme progressed. 2a - increased self-awareness and insight, greater personal resilience and improved motivation. 2b - participants "appear to have gained tremendous personal benefit" which focused on confidence to operate outside their initial comfort zone; greater clarity about their leadership role; greater assurance about their own leadership capability and sense of empowerment; greater awareness and knowledge of policy issues and the need to develop skills in the areas of strategic influence. 3a - improved relationship-building and networking skills and the development of improved influencing ability. 3b - positive feedback was also offered by those around the participants in their employing organisations. In responding to telephone interviews and questionnaires, these colleagues, managers and staff of participants reported significant behaviour change. 4b- successful completion of significant service development through leadership projects had made significant impact locally and nationally.	5	10.5
MacPhail, 2015 [37]	Single Australian Hospital Department (Geriatrics, Rehabilitation and Palliative Care)	39	Non-executive middle or senior level medical, nursing and allied health professionals	9-10 months	2-hour monthly sessions, group projects, site visits	Guest speakers and discussions external site visit, project, presentation	Organisational structure, healthcare context, leadership and patient safety, quality of care and clinical errors, complex systems, communication, engagement with patients, conflict, negotiation, change management	1. Participants reported almost unanimously (86-100%) for all measures including that the course was relevant and valuable 2a. Participants reported increased willingness to take on leadership roles 3a. Participants reported in comments that they benefited from increased networking and relationships across silos. 3b. 4 of 17 participants from the first cohort had been promoted 18 months after the programme (no control group available)	5	10.5
Cerrone, 2017 [38]	Large US integrated health care organization	80	Residents (Incoming Chief Residents)	1 day	9hrs	Emotional intelligence inventory three weeks prior to course didactic sessions simulated teaching encounters	Leadership, managing, and core feedback skills, interpretation of emotional intelligence inventory, interpersonal and communication skills and professionalism	1. 92% agreed that the program met their learning needs. Participants reported considering the OSTEs useful. 3b. OSTE scores (out of 100) increased from mean = 47.92, SD = 7.8 to mean = 51.22, SD = 6.9); t (68) = 1.99, p = 0.006	1	10.5
Patel, 2015 [39]	Single US hospital	30	Residents (PGY2-4)	2 years	2 year medical leadership in quality residency track	Core curriculum (120 hours over 3 weeks including lectures, readings, videos, small group activities, online modules, facilitated discussions), integration into a quality improvement leadership team, capstone qi project, mentorship	Methods and tools of quality improvement and patient safety, human factors engineering and safety culture.	1. The core curriculum has been rated as highly valuable and necessary (mean 4.95 and 4.96/5 respectively) 2b. Mean improvement of quality improvement knowledge assessment was 3 points for cohort 1 (SD 3), 4 points for cohort 2 (SD 1.6). Improvements were marked by the track directors. No significance reported 3a. Graduates reported that they intend to pursue quality/safety leadership positions (13 strongly agree, 1 agree) 3b. All graduates have completed their projects. 12 posters presented at national conferences regarding capstone qi projects. 4b. "most of the capstone projects have resulted in improvements in health care delivery" - not otherwise specified.	n/a	10.5
Nakanjako, 2015 [40]	Four African and 4 US universities	15	Unclear	1 year	1 year fellowship	8 weeks of didactic teaching, with two 4.5 month experiential trainings at health organisations. Also 4 online modules. Mentoring (weekly meetings, monthly mentoring team meeting). Logbook for learning.	Leadership, communication, monitoring and evaluation, health informatics, research methodology, grant writing, implementation science, and responsible conduct of research.	2b. Participants reported improved skills and knowledge 3a. Participants reported use of new knowledge and skills in both current and new leadership roles 3b. All graduates have remained in health leadership positions in Uganda. 86% (13/15) have opted to take on new responsibilities. 4b. Individual projects completed on the programme had a range of positive outcomes. These amounted primarily to official recommendations and initiation of new pathways, but several of the projects are reported to have improved patient care.	2	10
Kuo, 2010 [41]	Single US residency program	16	Residents (Paediatrics PGY1-3)	3 years	Bespoke residency programme	Small-group seminars, project work, and mentoring	Themes of leadership, critical thinking, and community engagement. Topics include policy making, project management, decision making and communication	1. Satisfaction with programme 3.73/4 ±0.46 2a. Positive impact on plans to influence population health and health policy 3.53±0.64 and to serve minority or underserved populations 3.47±0.74 2b. Improvement of competence as a leader 3.40±0.74 3b. Participants have received both local and national awards recognizing their leadership and commitment to the community. Nine graduates are in positions of leadership such as medical directorships 4b. 9/16 graduates have received grants to support their projects. Multiple projects have achieved sustainable funding and impact across advocacy, health programme development and policies.	1	10
Brandon, 2013 [42]	Single US residency program (radiology)	44	Residents (Radiology)	1 year	7x 90min modules	lectures and case-based group discussions	Finances, quality improvement, employment, organisational dynamics, healthcare policy and economics, negotiation and conflict management	2a. Significant improvement in participants' knowledge for all modules (p<.001) 2b. Significant improvement in participants' self-assessed confidence scores for all modules (p<.001)	n/a	10
Green, 2002 [43]	US network of community- owned health care providers and physicians	26	26 Teams from eight organizational units	2 years	Coaching and leadership initiative	Faculty coaching Quality improvement projects Team meetings, with team learning sessions and planning for six-month action period following the meetings. Teams from subsequent waves overlapped	Diffusion of innovation, Strategic goal setting, engaging others, PDSA, barrier-busting and infrastructure-building, project management, reflective thinking and learning, conceptual thinking, summarizing and communicating, coaching, and building further organizational capacity for spread	4b. 17 of 26 teams reported significant clinical improvements in targeted areas, improvement work has become easier with each cycle (improved from 50% of projects within 3 weeks in cycle 1, to 100% in cycle 3)	n/a	10
Hemmer, 2007 [44]	Single US residency/ fellowship program (pathology)	16	Residents and fellows (Pathology)	1 year	6 x 1-2 day workshops (average 10hrs per workshop)	Pre-reading, didactic lectures, interactive sessions, case	Leadership and management basics, managing change and interpersonal skills, personnel	1. Participants evaluated (five-point scale) the content and speakers (scores from 4.4 to 5.0).	n/a	10

						scenarios, team- building exercises, project	issues, quality, informatics, finance, and a capstone seminar	2b. Participants showed significant improvement in their leadership and management test scores (61/62 % to 88% in two different cohorts)		
McCurdy, 2004 [45]	Single US academic medical centre	22	Faculty members (who were, at the time of the course, in a leadership position or likely to move into a leadership position soon)	12 months	Eight 3-hour sessions in two 2-day blocks 6 months apart and 2 evening small group discussions	Workshops, evening sessions, Project with poster or a short oral presentation	Setting goals, leadership, change, emotional intelligence, organisational dynamics, mediation and negotiation, quality improvement	2b. Post-program self-assessments significantly and meaningfully different from pre-programme using both traditional and retrospective measurement. 3b. 14 projects completed	n/a	10
Hadley, 2014 [46]	Single UK training deanery	30	Residents (PGY2 doctor/s paired with a management trainee)	6-9 months	Paired quality improvement project with a management trainee and a PGY2 doctor	Project work, mentoring, action learning sets	No formal curriculum	1. Continued engagement from 24/30 participants for the programme 2a. Participants felt empowered to start service improvement projects 2b. Statistically significant increase in self-reported understanding of several domains (statistics and data not reported) 3b. Projects presented by 9 of 14 teams	3	9.5
Revere, 2015 [47]	Single US hospital network	50 "approximately"	Senior physicians "chairs and near-chairs"	6 months	Fortnightly full-day sessions, project work	Lectures, discussions, projects	Strategy, finance and accounting, organizational performance measures, including clinical quality, human resources and customer experience management, organizational behaviour and team building and leadership skills.	1. Course rated 4.69/5. 71% rated topics as relevant 2a. Participants reported changed perspectives 2b. Participants reported learning from the lectures 3a. Participants reported being more engaged in their conversations, and applying learning day-to-day 4b. 40 per cent of the course projects have been funded by and implemented within the participant's respective institution. Senior sponsors have continued to be satisfied with project and programme outcomes and continue to send participants	3	9.5
Osborn, 2004 [48]	Ten US Paediatric Association locations	32 (total 2 cohorts)	Paediatricians from AMCs	2-3 years	Three 2-day workshops during first 12 months, then 2 advanced workshops annually (only one for cohort 2)	Workshops, project	3 Clusters: 1) institutional leadership concepts; institutional leadership skills; management skills (very good table in the paper with details of each).	2b. Reported increased knowledge and skills in all areas related to giving workshops 3a. Participants reported that they incorporated into their workshops knowledge and skills gained in 1) mission, vision, and values; 2) how to approach organisational change; 3) mission-based management; 4) net- working and team building; and 5) workshop development 3b. Across a 2-year follow-up period: 13/30 (43%) conducted a workshop, 2/30 (7%) were promoted, 18/30 (60%) participants assumed a leadership position participants conducted 57 local workshops and 33 regional or national workshops that were attended by 1082 participants. 4b. Across a 2-year follow-up period 8/30 (27%) obtained a peer-reviewed grant	2	9.5
Wichman, 2009 [49]	Single US hospital department	6	Residents (Psychiatry PGY3)	8 months	Weekly seminars (8 modules x 4 sessions each), with projects and mentoring	Lectures, discussions, projects, mentoring	Financial management, Human resources management, Planning and marketing, Information management, Risk management, Governance and organizational dynamics, Business and clinical operations, Professional responsibility.	1. All respondents said course met or exceeded their expectations. 3b. Multiple projects completed during programme 4b. One project resulted in decreased non-attendance by new patients by 50% across 12 months	n/a	9.5
Monaghan, 2018 [50]	Single UK hospital trust	12 (6 doctors)	Residents (non-training clinical development and clinical teaching fellows, managers of various departments)	6 months	6 month paired learning doctors/managers	6-month paired learning matching doctors with managers. Shadowing, conversations, reflections	Nr	2a. Physician participants reported feeling more prepared for a range of leadership requirements, including understanding decisions and working in teams and with managers 4b. Two pairs collaborated to successfully implement a lasting organisational change in the form of a trainee management forum	4	9
Voogt, 2016 [51]	Six Dutch teaching hospitals	NR (~50)	Residents (mixed specialties and levels)	1 year	Four 1hr meetings to discuss and plan projects	Facilitated discussions, project work	Quality improvement and leadership, not otherwise specified	2a. Interviewees reported feeling empowered 2b. Interviewees reported increased awareness of organisational aspects of healthcare delivery. 4a. Outcomes of projects for example, Development of new handover guidelines to ensure the attending physician is present at handovers	3	9
Heitkamp, 2017 [52]	Single US hospital department	98	Residents (radiology)	NR (variable, journal club)	One hour lunchtime Journal club meeting every fortnight	Journal club, projects, mentoring, leadership role placement	Leadership (topics chosen by the group on an ad hoc basis)	2a. Participants felt better prepared for their careers 2b. Self-reported increased understanding of the business of radiology 3a. Self-reported improved communication, interaction with referring physicians, career development team building, group dynamics, and interactions with hospital administration 3b. 35% had assumed leadership roles within three years. Publication of 27 articles in core radiology journals, completion of multiple projects 4b. New external collaboration for residents	0	9
Pearson, 2018 [53]	UK leadership fellowship	12	Residents (senior, medicine, general practice, surgery, obstetrics and gynaecology, paediatric surgery and psychiatry.	1 year	1 year out-of-programme fellowship in a host organisation	Symposia and conferences, one-to-one coaching sessions, action learning sets, shadowing opportunities and reflective practice including completion of a portfolio. Project work for host organisations	Minimally reported. Communication, working styles and leadership framework mentioned	1. Most participants reported that all components of the course were very or slightly useful. 75% did not access multi-source feedback component. 2a. Most participants reported improved attitudes towards leadership and their ability to make changes in their organisation. Reports of increased confidence. 2b. Reports of increased awareness of other working styles and characteristics of good leadership. 3a. Reports of adjusting behaviour towards other working styles	0	9
Crites, 2004 [54]	Single US residency program (internal medicine and paediatrics)	13	Residents (PGY1-4)	1 year	Monthly seminar series	Interactive lectures	Coding, finances, group dynamics, human resources, risk management	1. Participants rated the course effectiveness 4.13/5 2a. The residents' views towards practice management education in general had mean scores of 4.67 (/5) 2b. Participants scored significantly higher on a self-assessed management skill, from 2.62 to 3.65 (/5). Average score on knowledge test significantly increased from 74% to 91%	n/a	9
Dickey, 2014 [55]	Single US hospital	Unclear (3 leadership projects described)	Residents (Psychiatry PGY1-4)	4 years	Modular leadership programme over 4 years with mostly voluntary components	Seminars, with voluntary simulations, action teams work, electives, mentoring	Philosophy of leadership, healthcare delivery systems, quality assurance, risk management, qualities of exceptional leaders	3b. Three projects undertaken as part of voluntary elective 4b. One project resulted in a reduction of admission time of 65 minutes	n/a	9

Foster, 2008 [56]	Single US medical centre	12	Residents	2 years	3 week intensive orientation, 11 month MPH degree and leadership coursework, weekly 1/2 day didactic sessions, monthly journal club, monthly open evening sessions	MPH degree, leadership coursework change project, mentoring	Leadership of small systems in health care Measurement of illness burden in individuals and populations Measurement of the outcomes of health service interventions, Leadership of change for improvement of quality, value, and safety of health care of individuals and populations Reflection on personal professional practice enabling personal and professional development.	3b. Projects completed 4b. 8 projects completed with patient and organisational benefits for example, Decreased procedure and hospitalisation times	n/a	9
Freeman, 2018 [57]	Single US training programme	30	Cardiology fellows-in-training and early career professionals	2 years	2 year cardiology leadership academy	Mentoring, not otherwise clear	Conflict management, team leadership, influencing others, navigating challenging conversations, and how to achieve work-life balance.	2b. Participants reported increased confidence across a range of key curriculum areas 3a. Participants reported the programme had helped them to gain new leadership roles, and that they had applied learning from the programme in their new leadership roles; 4a. Of those that acquired a new leadership opportunity, 100% of respondents reported participating in the leadership academy program had an impact on their success within the newly acquired leadership role(s).	n/a	9
Saravo, 2017 [58]	Single German university hospital	50	Residents (PGY1-4 across specialties)	4 weeks	Weekly 2.5hr sessions after clinical duties (x4)	Didactic module, standardised simulations, one-on-one feedback on recorded simulations, "practicing communication techniques"	"Full Range Leadership Model" (Bass); transactional and transformational leadership, simulation of critical incidents, communication techniques	2a. No change in knowledge in exposure or control. No difference in self-assessed change in self-assessed leadership scores compared with control (both had small increases in mean scores). 3b. Increase in observed performance on bespoke transactional and transformational scales (based on video recorded simulations, no control group)	n/a	9
Schulz, 2013 [59]	US surgical department specialty	9	Residents (ENT)	6 months	Virtue strength assessments (VIA); mentorship meetings; 'thought of the day'; internal and external faculty training and development; leadership basic training course	Mentoring, interactive sessions mixed with reading, individual character assessment	Virtue Leadership, curriculum not otherwise specified	1. 100% agree that project was a valuable learning experience (up from 56%) 2a. No significant change in attitude towards leadership 2b. Increased knowledge of the leadership values of the organisation (94% up from 47%)	n/a	9
Stoller, 2007 [60]	Single US hospital network	Unclear (roughly 400)	Emerging physician-leaders	9 months	Eight sessions offered roughly once monthly on Fridays at an off-site retreat centre, Longitudinal project	Seminars, business case project	Marketing in healthcare, Healthcare finance, Writing a business plan, Emotional intelligence, situational leadership, conflict resolution and negotiation, Medicolegal issues	3b. 49 business plans were submitted over 13 courses. 4b. 30 of 49 business plans have been implemented	n/a	9
Wulfert, 2017 [61]	German Master's Degree (Leadership in Medicine)	17	Varied (2 senior consultants, 8 consultants, two senior residents, one pharmacist, one financial controller)	>100 days teaching time over a master's degree	Master's degree	Project work, didactic teaching, discussions, thesis	Leadership and management, quality management and legal aspects, medical ethics, project management, empirical research methods, communication/dialogue skills, change management, diversity management, management accounting, health economics, medical engineering and computer science, education and didactics, educational management	1. 14/17 participants completed the masters in full 3b. More than 30 projects implemented over the course of the masters. 4b. Projects successfully integrated into existing structures	n/a	9
Blumenthal, 2014 [62]	Single US hospital	16	Residents (PGY2 Internal Med)	4 weeks	2-3hrs a week for 4 weeks, work between	Large-group discussions case-studies, videos, role-plays small-group meetings homework/required readings	Clinical leadership Goleman's leadership types and emotional intelligence authentic leadership effective team leadership	1. Sessions perceived to be relevant 2a. Increased confidence, feeling more prepared for team challenges 2b. Self-reported learning, better understanding of personal strengths and weaknesses as a leader	4	8.5
Sanfey, 2011 [63]	Single US university	143	Academic faculty members who had demonstrated leadership potential within the school of medicine.	10 weeks	3 to 7 hour weekly training sessions over a 10-week period for an approximate total of 40 hours.	Instructional sessions, self-analysis using MBTI and the leadership skills inventory 360	Managing organizational change, making strategic decisions, assessing the dynamics of successful leaders, financial management, and finding life balance in a growing career.	Immediately post: 2a. 50% of participants indicated that their perspectives on leadership had changed as a result of the programme. 2b. For each leadership skill surveyed, the majority of participants agreed that it was improved by programme participation 3a. 81% of participants indicated that they had changed their professional behaviour as a result of the programme. 3b. 66% of respondents were invited to apply for or offered positions post-programme. 1-4 years post: 2a. Some respondents reported increased insight into others/self/job/leadership. 2b. Respondents reported increased ability to recognize weaknesses and strengths, and a range of other increased leadership skills. 3a. Only 27% of respondents reported changes in their behaviour, though 31% reported that they had been very successful in achieving their career goals.	2	8.5
Edler, 2010 [64]	Single US residency programme (paediatric anaesthetics)	Unclear	Residents (first year paediatric anaesthetics residents)	1 year	1 year administrative resident programme	Reading, experiential learning, feedback, self-assessment, mentoring	Organisation culture, human factors, quality assurance (QA) and continuous quality improvement (CQI), operating room scheduling, and resident selection, decision making, technical planning, interpersonal or professional actions, and conflict resolution	1. Qualitative report of satisfaction with the programme 2b. Report of improved conflict-management 3b. Observed application of leadership skills in programme improvement	1	8.5
Richman, 2001 [65]	US/Canada executive education in academic medicine	200	Midcareer female faculty at medical or dental schools at associate or full professor rank	1 year	Two week-long residential sessions (in September and April), annual conference,	Lectures, panel discussions case studies, computer simulations, role playing, small group work, individual interviews and projects,	"mini-MBA, contemporary leadership issues, personal professional development (individual assessment tools, conflict management and negotiation skills, team-building skills)	2a. Reports of improved confidence in addressing and resolving conflict situations. Insight and confidence into "how the game is played". 2b. Reports of improved knowledge in addressing and resolving conflict situations. Assessment showed significant and large increase in all curricular areas. (p<0.0001) reports of increased knowledge of career possibilities, new insights into how to advance their own careers, and understanding of a greater range of options open for their pursuit.	0	8.5

					numerous assignments	extensive individual assessment, coaching		Increased awareness of educational and medical issues and development of strategies to be informed, resolve problems, and advance projects. Understanding of new leadership and management strategies, allowing fellows to conceptualise, introduce, and implement ideas with greater effectiveness and confidence 3b. Fellows have been successful in advancing to higher leadership roles		
Farver, 2016 [66]	Single US hospital	105	Residents (New Chief Residents)	2 days	2 day Chief Residents' Leadership Workshop	Pre-workshop readings Workshop Post-workshop readings	Teambuilding Conflict Resolution Negotiation Emotional Intelligence Physician Health Mentoring	1. All workshops rated highly 2a. Confidence in team building increased 2b. Increased perceived ability to locate areas of need. Increased familiarity with workshop concepts. 3a. Self-reported increased ability to resolve conflict effectively	n/a	8.5
Gregg, 2016 [67]	Single US trauma centre	20	Residents (Trauma, y3-5)	6-30 months	Evaluation of communication skills and weekly discussion at meetings	Evaluation and feedback	Communication	3b. Increase in observed communication/professionalism, systems-based practice, medical knowledge, practice-based learning, patient care	n/a	8.5
Hill, 2018 [68]	Single US hospital	7	Residents (senior surgery)	3 weeks	3 week course with senior residents giving 5 presentations a week to juniors. Timing and frequency not specified	Presentations by participants to junior residents reflecting on readings from the course reading book	Weekly readings from the book "the founding fathers on leadership"	3a. No statistically significant changes in survey results relating to behaviour with Bonferroni correction applied.	n/a	8.5
Pugno, 2002 [69]	US residency director program	"more than 300"	Residency directors (family practice). Numbers not specified	9 months	A three-day conference and two one-day sessions. Project work, mentoring	Didactic and small-group sessions, project	Leadership skills, personnel management and team building, program management, communication skills, negotiation skills, program finances, educational guidelines for family practice	1. 99% of participants rated the programme valuable (15%) or very valuable (84%); 2a. Enhanced job satisfaction, reduced job stress, and an expanded network of educational contacts and resources. 4a. 76 % reported that the program lowered the level of stress. Participants reported enhanced job satisfaction, reduced job stress, and an expanded network of educational contacts and resources. 83% said they were more likely to continue as a program director for the next few years after the programme. The average tenure of family practice program directors has increased from 3 to 4 years before the programme, to more than 6 years after the programme had been running for 5 years.	n/a	8.5
Denney, 2019 [70]	Single UK deanery (South East Scotland)	89	Residents General Practice Specialty Trainee, Year 1)	6 months	Single session on leadership, recommended menu of possible experiences	Guidance, formative feedback, encouragement to seek out specific leadership activities	Recommendations: chairing a meeting, "fresh pair of eyes" exercise NOS, running an educational session, practice leaflet project, clinical protocol, website design, mini-quality improvement project	2a. All trainees who completed the survey reported that development of leadership skills is either "quite important" or "very important" for a trainee's future career. 51% of trainees felt more involved in their GP practice as a result of their leadership activity. 2b. There was no significant change in the self-rated confidence in leadership skills or team working skills when compared pre-post. 3a. 83% of respondents reported having undertaken a voluntary leadership activity, particularly quality improvement projects and running educational sessions. Those who did not, frequently reported time and service delivery pressures as a reason.	5	8
McAlearney, 2005 [71]	Single US hospital	52	Physicians (two cohorts)	2 years	1hr monthly morning sessions, 1 half day session each 6 months	Adult learning. Interactive questions, case-based scenarios.	Decision making, problem solving, strategy, change, situational leadership, communication, negotiation, coaching and mentoring, conflict, finances	1. Mean score (/5) 4.7 for satisfaction, 4.8 for quality of information 2a. Mean score (/5) for perceived applicability 4.6. Mean confidence in strategic direction of the organisation 4.1 2b. Participants reported increased awareness of leadership resources (4.1), motivation to be involved in their community (3.7) 3a. Participants self-reported increased leadership effectiveness (4.2/5), teamwork (4.0) and team leading (4.3), as well as new roles (4.0). Participants reported using skills learned; 4a. Improved organisational collaboration towards strategic initiatives	5	8
Shah, 2013 [72]	Single UK specialty training	40	Consultant ophthalmic surgeons	2 days	2x full day interactive sessions	Gaming, team challenges, meta-planning, role play and professional actors, interactive presentations, and self-analysis	Admitting vulnerability and uncertainty, taking responsibility for managing risk, being self-aware and reflexive, internalising authentic leadership	2a. Participants reported increased self-awareness 2b. Demonstrated increased knowledge of authentic leadership concepts 3a. Self-reported long term (4 years) application of skills learned into clinical practice 4a. Participants reported improved performance for them and their teams as a result of the programme	2	8
Clapp, 2018 [73]	Single US hospital department	36	Residents and faculty (Department of Anaesthetic and Critical Care)	1 year (presumed)	2x 2hr evening skills session, 1 half-day session and 1 2.5hr evening session "capstone"	Capstone: presentation on relevant topic small-group discussions presentations back to large group	Feedback supporting colleagues during tough times clinical pedagogy	1. 31% excellent, 44% very good, 25% good; 4a. "the frankness of the discussions between residents and attendings in these sessions has paved the way for a broader departmental initiative designed to facilitate more effective leadership and teamwork among individuals in all departmental roles—attendings, residents, nurse anaesthetists, administrative staff—through identifying the strengths and deficiencies of the ways in which they interact.	1	8
Pettit, 2011 [74]	Single US hospital department	11 (9 responses)	Residents (Neurosurgery)	1 year (academic)	Monthly 1hr sessions	Interactive lectures, self-assessment activities, case studies, self-reflection, discussions, and reading materials	Leadership style, conflict management, effective feedback, team building, team leadership, motivation, moving from peer to leader.	2a. Qualitative comments indicate more open attitude towards leadership roles 2b. Significant self-assessed increase in leadership knowledge.	1	8
Donnelly, 2016 [75]	Single US residency programme	Unclear	Residents (radiology, PGY1-4)	4 years	4 year residency programme in leadership fundamentals and leadership tracks.	Monthly lectures for first 2 years, research work with imaging scientists, mentoring, project work	Research, education, business/management, quality care/service, and information technology	1. "universal praise" from participants for the mandatory first 2 years, 100% opted into the optional years 3-4. 3b. 4 participants have worked on projects as part of the programme, with one having submitted academic manuscripts resulting from her project	0	8
Gruver, 2006 [76]	Single US health system	17	Physicians and non-physicians (finance, IT, nursing, public relations)	8 months	Monthly seminars over dinner (2-2.5hrs)	Case-based leadership discussions during two-hour sessions	Managing vs. Leading, forming a vision, predefining a person's moral compass, risk-taking and transactional leadership	1. High ratings for the course 4.63/5 for comparison to other leadership and education programmes experienced. 2a. Improved self-confidence in leadership, intended changes to leadership style 2b. Increased knowledge of desirable leadership characteristics in the organisation 3a. Several of the participants reported experimenting in their current leadership	0	8

								assignments with concepts discussed during a session; one participant used one of the cases to heighten the leadership awareness of some of his own subordinates. Reported personal bonds of participants with the CEO.		
Babitch, 2006 [77]	Single US residency program (paediatrics)	NR	Residents (Paediatrics PGY1-3).	9 months	Nine sessions	Lectures	A core curriculum focusing on physician compensation, medical economics, healthcare system, leadership and communication, career/CVs, contracts, health law, and customer service	1. Satisfaction scores “between 3 and 4” on a four-point scale. 2b. Improvement in tested comprehension of the subject matter of each lecture, with an average increase of 20% to 40% between tests (5-point scale)	n/a	8
Gulati K, 2019 [78]	Indian health care organisation	96	Clinicians and hospital administrators from public and private sector organisations	3 days	3 day leadership residential programme as part of a 6 day residential programme	Didactic lectures, small group workshops, focus group discussions, case-based discussions and experience sharing.	Leadership styles, leadership competencies, team building, strategic management, procurement, operational excellence, legal and ethical issues, budgeting, financial management, conflict management, quality and patient safety, hospital accreditation, communication, human resource development, health-care technology, contract management, hospital projects and supply chain management.	2b. Statistically significant increase in all 30 items related to knowledge and skills	n/a	8
Stoller, 2004 [79]	Single US residency program (internal medicine)	32	Residents (PGY-1)	1 day	One day retreat	Group simulation exercise, group discussion	Team skills, group dynamics, leadership	1. All attendees rated the retreat as valuable. 2a. Based on significant changes in residents’ responses on the post retreat questionnaire attendees believed that the retreat enhanced their abilities to be better physicians, resident supervisors, and leaders. (all p<0.001) post retreat responses (table 3) indicated significant increases in agreement that good leaders challenge the process, make decisions based on shared vision, allow others to act, recognize individual contributions, and serve as good role models.	n/a	8
Edmonstone, 2009 [80]	Multiple UK strategic health authorities	200 (approx.)	Senior medical leaders in primary and secondary care and public health	12 months	Two 3-day residential modules at the beginning and end of the programme. Three interim event days.	Personal development plan, coaching, mentoring	Leadership for partnership Personal development – through the creation of a personal development plan, provision of coaching, mentoring, etc. Working in networks Lateral thinking/scenario planning Leading change	1. Participants dissatisfied with the programmes, partly due to high expectations not being met. Falloff in attendance 2a growing sense of identity as clinical leaders. Increased confidence. 2b. Greater appreciation of others' roles and perspectives 3a. Tools and skills applied in practice, increase in personal networking. 3b. Projects completed as part of the programmes.	5	7.5
Stergiopolous, 2009 [81]	Single Canadian residency program	52	Residents (PGY2=24, PGY4=28)	4 half days	Workshops (four half-days)	Interactive teaching as much as possible. Didactic teaching and small groups or other interactive techniques (buzz groups, brain- storming, think-pair-share discussions, a debate, and clinical case studies)	Teamwork, conflict resolution, quality improvement, program planning and evaluation, leadership and change management, mental health reform, organizational structures, and self and career development	1. Attendance averaged 54% overall. Workshops rated average of 4.2/5. Participants appreciated the reflective and interactive components of the workshops and valued the hands-on exercises and the use of case studies and “real life” examples. They suggested that more time be dedicated to quality improvement and medical error and opportunities to take part in administrative committees and quality improvement projects at their hospital sites. Focus requested on current efforts rather than historical overviews. 2b. Objectives related to skills and knowledge rated as met (~4/5 on Likert scale)	2	7.5
Berkenbosch, 2014 [82]	Single Dutch university medical centre	14	Residents (O&G, orthopaedics, paediatrics, internal medicine)	8 hours	2 x 4 hour sessions 3 weeks apart with homework between	Didactic teaching/lectures student presentations simulation	Knowledge of the healthcare system time management	1. Rated 7.66/10, "it fills a gap in our current postgraduate medical training" 2a. Increased interest in leadership development 2b. No significant changes (underpowered)	1	7.5
Block, 2007 [83]	11 residency programs in Australia	146	Residents (134 registrars and 12 resident medical officers)	2 days	2 day programme	Experiential small-group work, individual exercises, self-analysis questionnaires, videos, simulations, didactic content	Leadership competencies, self-awareness, communication and learning styles, conflict resolution, serving as teacher, time management, delegation, leadership styles, managing stress, safety and quality, team building, feedback and action planning	1. High satisfaction scores of 6.2 on a scale of 1 to 7 (sd0.6) 2a. Reported desire to explore leadership and management in greater depth. Increased awareness that the transition to being registrar was more multifaceted than it may seem.	0	7.5
Donaghy, 2018 [84]	Single UK hospital trust	NR (>140)	Specialty trainees ST4-ST8	10 months	14 sessions 90-120 minutes	Large group lectures discussion and reflection action learning (qi or patient safety) trainee presentations at workshops	Medical leadership project management patient safety and QI methodology trust overview and patient client experience high performing teams human factors interview preparation clinical networks commissioning	1. 100% of respondents would recommend step to a friend 2a. Improved confidence, intention to apply 2b. Improved understanding in a range of domains 3b. Completion of some qi projects	n/a	7.5
Frugé, 2010 [85]	Single US hospital department	39	Residents (Paediatric Oncology)	NR	Twice-monthly seminar series	Guided reflection on challenging leadership experiences	Leadership reasoning	1. 75% of fellows continue to participate in later years despite no longer being mandatory. 2a. Significant increase in confidence in all items on the bespoke questionnaire. Examples include working effectively as a team, effective communication, self-awareness.	n/a	7.5
Hunt, 2017 [86]	Single US training camp	30	Residents (Interns at the start of PGY1 surgery)	NR	Surgical pgy1 boot camp	Didactic material, narrated lectures, procedural videos, course manuals. Personality assessment. Non-leadership specific simulation	Unclear	1. Participants agreed that the information was relevant to their future careers 2a. No significant increase in self-awareness (only 55% agree or strongly agree that understanding of own behaviours and motivators improved immediately post, only 40% agree or strongly agree at 3 months)	n/a	7.5
Murdock, 2011 [87]	Programme across 3 US states	>100 (not specified - 'five cohorts of 20 or more')	Community practice physicians (five cohorts)	20-weeks	Weekly three-hour evening sessions	Weekly three-hour evening sessions	The business of medicine, quality improvement, transformational leadership	2b. Increase in self-assessed competency in all the 26 categories in each of the program’s five cohorts (significance not reported); 3a. Commentaries and assessments revealed an increasing level of empowerment in their leadership roles and increased desire for selection to leadership roles.	n/a	7.5
Thakur, 2018 [88]	Single UK hospital trust	23	Residents (PGY4-5 Psychiatry residents)	1 year	Six full day workshops;	Case studies, role-play, exercises, reading, mentoring	Medical leadership competency framework (demonstrating personal qualities, working with	1. All workshops rated between 3-4/4 2a. Participants felt the programme helped them in achieving leadership competencies. 2b. Residents described how the conceptual issues learned in the workshops were	5	7

					mentoring programme		others, managing services, improving services, setting direction)	reinforced with leadership champions and then applied to their projects 3b. Participants completed and presented projects at local medical education day.		
Patterson, 2013 [89]	Single UK GP deanery (South Yorkshire Region)	8	Residents (GP trainees, PGY3)	8 months	Facilitated leadership projects	Project work, personal reflections, facilitated monthly discussions	Leadership, change management, and teamwork skills	1. Participants reported having enjoyed the programme 2a. Participants reported being inspired, feeling more mature as doctors. 2b. Increased self-awareness. In many domains there was a negative shift in self-assessment by the end of the programme - thought to be "re-calibration of confidence", supported by focus groups. Focus groups also reported learning around the difficulty of change management. 4: one participant was noted to have achieved change implementation through their project (not otherwise specified)	4	7
Maza, 2016 [90]	Single Israeli health provider	256	Physician-managers	8 weeks	5 full days over 2 weeks with one overnight, followup meetings at 3 and 6 weeks	Theoretical knowledge, experiential learning, practical tools, deep personal exercises and simulations. Individual, dyadic, and group learning	Models of self-awareness, outcome thinking, determining a personal and organizational vision, and creating a personal approach to	1. Mean rating of 5.7/6 (post) and 5.4/6 (retrospective) 2a. Respondents reported increased self-awareness, which was sustained at 6 months. (>5/6) 2b. Respondents reported increased personal leadership ability, sustained at 6 months. (5.2/6 then 4.9/6) 3a. Respondents reported increased proactivity in management.	3	7
Steinert, 2003 [91]	Single Canadian department	16	Faculty (family medicine)	2 days	Two-day workshop	Interactive modules and exercises	Time management, goals and priorities, leadership styles and skills, and conducting effective meetings	1. All participants rated workshop as “very useful”. 2a. Several the participants reported that they would change their behaviour after the workshop, regarding time management, goals and meetings. 3a. Most respondents had successfully attempted determining short-term goals; handling paper more effectively; determining their ‘prime time’; protecting time for specific tasks; and setting meeting agendas. They were less successful at: delegating; saying ‘no’; adopting different leadership styles; and evaluating meetings.	3	7
Satiani, 2014 [92]	Single US hospital department	24	Surgeons who wish to assume administrative or leadership roles	18 months	4hr seminar once a month	Seminars, project work (teams)	Leadership competency, strategic planning and vision, financial management, business planning, communication skills, change management, quality of care and patient satisfaction, teambuilding, negotiation and problem-solving, stress/burnout and lifestyles issues, human resources and talent management, diversity for healthcare leaders, healthcare law, medical ethics	1. 100% of respondents would recommend the programme to peers 2a. Respondents reported increased confidence in leadership roles 2b. Respondents reported increased leadership skills, knowledge and self- awareness. 3a. Respondents reported using skills from the programme in their practice and having made networks they otherwise would not have made.	2	7
Lee, 2004 [93]	Single US residency programme (paediatrics, Hawaii)	10	Residents (PYG2)	3hrs	3-hour interactive workshop during resident retreat	Case scenarios, problem solving, role playing, interactive discussions and self-reflection	Managing teams, leading residents, and working with different personalities.	1.written comments from workshop evaluation forms were "overwhelmingly positive". 2a. Resident confidence increased significantly for managing teams ($p< 0.015$), leading junior residents ($p< 0.005$) and leading group discussions ($p< 0.017$). 2b. Self-assessment of leadership skills also increased significantly ($p< 0.043$)	0	7
Steiner, 2004 [94]	Single US department of psychiatry	13	Residents and post-doctoral fellows	5 months	8 x 1.5h seminars	Seminars with guest speakers and case discussions	Exploration of leadership roles in research, clinical practice, teaching, and administration; organizational dynamics and gender; negotiation skills and conflict resolution strategies; role of consultation, seminars, peer support, and mentoring by both men and women in the development of leadership skills; “keeping the balance: work, relationships, and personal health.”	1. Participants believed the course should be offered again and had a positive effect on their professional lives. 3a. One participant said it encouraged her to seek out mentors	0	7
Awad, 2004 [95]	Single US residency program (surgery)	NR	Residents (Surgical)	6 months	Nr	Not specified	A “focused program” to train residents to have the capacity/ability to create and manage powerful teams through alignment, communication, and integrity	2b. Statistically significant increase in score on a 34-item internal strength scorecard: alignment +13%, communication +12%, and integrity +12%.	n/a	7
Bayard, 2003 [96]	Single US residency program (family medicine)	NR	Residents (Family Medicine PGY2-3).	2 years	Nine half day sessions (first year), monthly one-hour sessions (second year)	Interactive-lectures and group assignments	A practice management curriculum: Determining/ balancing personal and professional goals, practice opportunities, facilities, organization, operation and management. Staff policies, legal issues, marketing, resources and hospital issues	1. Reported that the course was beneficial, a positive experience. 2a. Reported increased interest (and knowledge) in practice management 2b. Self-reported knowledge/comp fit with each of 13 practice management topics increased by roughly 2 points (five-point Likert scale)	n/a	7
Hanna, 2012 [97]	Single Canadian hospital	43	Residents (PGY3-6 Surgery)	1 day	One-day conference	Interactive lectures and case-based discussions; role-play, scenario simulation, small-group problem-solving sessions, and live feedback sessions.	Giving feedback and delegating duties, building teamwork, managing time, making rounds, coping with stress, effective learning while on duty, teaching at bedside and in the OR, and managing conflicts. Negotiating employment, managing personal finances, hedging malpractice risk, and managing a private practice	1. 79% felt that management was well-addressed or very well-addressed cf. 5% before the course 2a. Statistically significant improvement on perceived preparedness for all 4 managerial duties (negotiating employment, managing personal finances, hedging malpractice risk, managing a private practice) 2b. Participants reported improved skills in giving feedback, delegating duties, coping with stress, effective learning, and effective teaching. These were not significant with Bonferroni correction applied.	n/a	7
Patel, 2019 [98]	Single US hospital residency	17	Residents (PGY1-2)	2 years	Quarterly mentoring meetings for 2 years as part of a medical leadership programme	Mentoring	Leadership NOS	1. Participants rated their likelihood of recommending the programme at 7.8/10 (10 being extremely likely); 2a. Confidence seemed to increase (limited data reported) 2b. Gained confidence in time management, negotiation and leadership skills	n/a	7
Stefan, 2011 [99]	Single US hospital	24	Residents (Senior)	4 weeks	Weekly 1hr simulation sessions	Lecture, briefing, simulations, reflections	Advanced cardiovascular life support leadership, equipment management, and cardiac rhythm recognition and management.	1. "more than half of the participants considered the program to be beneficial"; 2a. Overall self-confidence score improved from 2.8 to 3.9	n/a	7
Gagliano, 2010 [100]	Single US hospital	90	Physicians with leadership responsibilities	2 years	Monthly sessions of 2–4 hours	Lectures and case-based discussion	Organizational leadership, financial management, management strategy, applied skills and tools	1. Participants reported high satisfaction with the course, and most would recommend to colleagues. 2a. Participants reported feeling more interested in and prepared for leadership responsibilities. 3a. 79% of participants reported having changed their approaches to projects or problems as a result of the course.	2	6.5

Gurrera, 2014 [101]	Single US hospital	8	Residents	5 months	1hr per week sessions for 5 months; team project	Didactic teaching, team projects	(1) overview of course; explanation of business plan project and expectations; resident interest survey; (2) organization and leadership models; (3) strategic planning; (4) the learning organization; (5) leadership/organizational ethics; (6) risk management; (7) marketing—part 1; (8) marketing—part 2; (9) decision-making; (10) lean management—part 1; (11) lean management—part 2; (12) QA processes and measures; (13) business plan workshop; (14) healthcare delivery systems; (15) microeconomics/ accounting; (16) medical errors/safety—part 1; (17) medical errors/ safety—part 2; (18) presentation of business plans.	1. Participants enjoyed the course and found it interesting 2b. Self-reported new knowledge. Self-reported reassessment of own personal skills 3a. Self-reported improved interpersonal skills and decision making	0	6.5
Bhatia, 2015 [102]	Single US institution	20	Residents (Internal medicine, surgery, emergency, >PGY2)	1 week	1 week programme	Case-based learning interactive talks small-group sessions simulation	Team dynamics leading change business of medicine communication skills	1. All topics rated >8/10	n/a	6.5
Hadley, 2015 [103]	Single UK deanery	NR (549 forms analysed)	Residents (FY2 doctors)	Single brief intervention	Leadership assessment and feedback	Evaluation and feedback	Leadership (personal qualities) Effective services (managing services), acting in a team (working with others), direction setting enabling improvement (improving services, reflection)	2b. 60% of participants felt that their leadership skills had improved as a result of the feedback received	n/a	6.5
Kasuya, 2001 [104]	Single US residency program (internal medicine)	NR	Residents (PGY1).	1 day	Six-hour retreat	Lectures and small- group tasks and discussions, scenarios and role play	Setting personal vision, leadership vs. management, building a team, practical negotiation skills, providing effective feedback, and problem-solving as a team leader	2a. Increased confidence in their abilities to lead a ward team (p = .0002) and fulfil their responsibilities as upper- level residents (p = .0002) and felt better prepared to deal with the challenges of being upper-level residents (mean = 3.65, SD .61). The participants also believed that they would use what they learned at this retreat as upper-level residents (mean = 3.88, SD .33). 2b. Reported having identified qualities they aspired to as upper-level residents (p = .0014). They also reported that as a result of the retreat they better appreciated their roles as team leader and manager (mean =3.76, SD .44)	n/a	6.5
Ninan, 2018 [105]	Single US Residency programme	NR	Residents (Anaesthesia)	1 year	Monthly 1hr didactic sessions	Didactic teaching, assignment	Personal branding, curriculum vitae, marketing, networking, evaluating and evaluating different types of medical practice, medical staff structure, governance, healthcare reform, future trends in medicine	Not possible to reliably infer due to poor quality reporting. 1. Junior residents expressed frustration that the program was taking them away from their clinical studies. Senior residents did not express this. 2a. Possible increase in knowledge across course objectives (no statistical analysis) 3a. Reported increase in quality improvement project involvement, scholarly production and networking behaviours, though the methods for this were not reported	n/a	6.5
Schwartz, 2014 [106]	International (US and Canada) Psychiatry leadership conference	541	Residents (all US and Canadian residency programmes)	3 days	3-day immersion course	Large and small group sessions, group tasks, peer and teacher feedback	Psychological challenges in leadership situations, personal conflicts, self-reflection and self-awareness, group process, conflict resolution, navigation of challenging leadership roles	1. Respondents found the feedback they had received to have been helpful (89%) 2a. Respondents reported improved understanding of group process and self-awareness 2b. Respondents reported improved leadership confidence and willingness to use conflict resolution skills and increased interest in pursuing leadership roles.	n/a	6.5
Ennis-Cole, 2018 [107]	Single US hospital	10	Physicians NOS	6 months	2hrs once a fortnight for 6 months (24hrs total)	Multi-source feedback, "insights" self-assessment and professional debrief, mentoring, classroom session, book club, reading, journaling, self-coaching, executive coaching edit: instructor led classes, assigned readings, self-directed learning via a binder resource guide, case studies and online experiences	Communication, development and learning, management and planning, relationship and team building, innovation and change, and patient centricity.	1. The attendance rate was 86% and the graduation rate was 96%. 2b. Participants reported increased team-building skills 3a. Participants reported daily implementation of skills learned	3	6
Steinhardt, 2015 [108]	Single US hospital	NR	Residents (GPGY4 Obstetrics and Gynaecology)	2.5 hours	2.5 hour workshop delivered by midwives	Roleplay, reflection, discussion, games	Principles of leadership, innovation, creative problem solving, and communication techniques	1. "nearly 100% numerical rating of 5" 2b. Participants reported improved insight into self and team 3a. Participants reported increased ability to be confident and vulnerable as a leader	3	6
Torbeck, 2018 [109]	Single US academic department	Unclear	Academic Faculty (New faculty up to senior leadership/chairs)	3 to 12 months	Four related programs for respectively new faculty, junior faculty, junior leadership, senior leadership.	Action learning, small group discussions, case scenarios, assignments, reflection exercises, multisource feedback, executive coaching (depending on which tier of the programme)	Leadership, communication, strategic planning, negotiation and conflict management, marketing, change, creating vision, managing difficult people, emotional intelligence, finances	1. Only 20% of faculty attended basic programme. Overall reaction post-session was positive to extremely positive. 3a. Participants noted in assignments that they had applied learning from the programme into their day to day work. 3b. Multiple participants interviewed for a leadership position.	2	6
Vimr, 2013 [110]	Single Canadian hospital	29	Physician leaders.	8 months	Five 1.5 day meetings over 8 months	Multi-source feedback, self-reflection, readings, action learning projects, coaching	Alignment of competencies, a systems and collaborative approach, affective learning strategies	1. "average rating for all components was 4.64 on a 6.0-point Likert scale". Components were not specified, nor the anchors of the Likert scale 2a. Comments cited increased awareness and understanding of leadership principles. 2b. Increased understanding of different leadership roles, accountabilities, and approaches.	2	6

								3b. Some of the individuals from the first cohort continued to work with their coach after the program was over.		
Bircher, 2013 [111]	Single UK deanery (extension of GP training)	NR	Residents (GP trainees)	2 years	Unspecified number of programme days	Didactic teaching, online learning environment, supervision, project work as individuals	The content was guided by the medical leadership competency framework, which includes domains of (1) delivering the service, (2) demonstrating personal qualities, (3) working with others, (4) managing services, (5) improving services, (6) setting direction	2a. Increased confidence in having difficult conversations 3a. Participants reported improved time management and application of skills learnt	0	6
Kochar, 2003 [112]	Single US academic medical centre	30	Faculty members	5 months	72hours - Nine-day course in three-day segments over five months	Sessions, lectures	Managing people, health care finance and accounting, leadership, marketing, health care informatics and information technology, health care quality, health care economics, time management	1. Overall course rated 4.6 out of 5 3b. Examples of cross-departmental collaboration have been seen.	0	6
Biese, 2011 [113]	Single US hospital department	NR	Residents (Senior Residents, Emergency Medicine)	1 year	Programme restructure to allocate leadership roles	Residents are allocated a leadership role in their final year of residency	Roles in: Administrative Chief, Resident Education, Resident Research, Journal Club, Medical Student Education, Ultrasonography Education, Resident Reading, Simulation, Information Technology	1. 100% of chiefs felt that the new system allowed the residency to meet more of its goals. All participants said they would recommend their position to other residents 2a. 100% felt their experience encouraged them to seek future leadership roles 3a. 100% felt their experience contributed to them deciding to enter academics.	n/a	6
Cherry, 2010 [114]	Single US university	141	Junior faculty	9 months	Two hours per week	Didactic classroom discussions, expert panel presentations, interactive case-based learning, group exercises, skill enhancement workshops, individual project with supervision	Setting goals, mentoring, negotiation and conflict resolution, performance review and compensation, presentation skills, facilitation, teaching, feedback, communication	1. Participants reported high satisfaction with the programs and with their mentor pairings 2b. Participants reported enhanced skills related to initiating and negotiating a new mentoring relationship 3b. "the project...often results in one or more scholarly products for the individual"	n/a	6
Johnson, 2014 [115]	Single US hospital department	Unclear (16 responses to survey)	Residents (Senior medical residents)	1/2 day	3hr seminar focused on emotional intelligence	Readings, formalized presentation, analysis of videos, role-play	Emotional intelligence	1. Seminar was felt by participants to have provided relevant content	n/a	6
O'Donnell, 2011 [116]	Single US hospital (residency programs)	NR	Residents (PGY1). Numbers not specified	4 weeks	Compulsory first year resident rotation in case management with 2hrs/week for 4 weeks	Lectures, discussions, case presentations	Overview of case management, advocacy, communication, and resource management	1. Evaluation broadly positive (>90% agreement with 6 statements); 2b. They acknowledge the case managers for their expertise, better understand utilization management, compliance, and coordination of care as a team and how this knowledge has also assisted them in their understanding of the continuum of care and regulations.	n/a	5.5
Ringdahl, 2014 [117]	Single US residency programme	36	Residents (Family medicine, PGY1-3)	NR	Unclear	Networking, mentoring, role play, simulations, discussions	Awareness of leadership opportunities, understanding organizational dynamics, conflict resolution, negotiation skills, mentoring, and personal wellness.	1. Feedback from participants has been uniformly positive	n/a	5

Table 17: Summary of included studies for the systematic review in Chapter Ten.
NR= Not Reported; NOS=Not Otherwise Specified; n/a=not applicable. MERSQI=Medical Education Research Study Quality Instrument; JBI=Joanna Briggs Institute Critical Appraisal Checklist for Qualitative Research.

References for Table 17: Summary of included studies for the systematic review in Chapter Ten.

1 Boyle DK, Kochinda C. Enhancing Collaborative Communication of Nurse and Physician Leadership in Two Intensive Care Units. *J Nurs Adm* 2004;34:60–70. doi:10.1097/00005110-200402000-00003

2 Parsons JR, Crichlow A, Ponnuru S, *et al.* Filling the Gap: Simulation-based Crisis Resource Management Training for Emergency Medicine Residents. *West J Emerg Med* 2018;19:205–10. doi:https://dx.doi.org/10.5811/westjem.2017.10.35284

3 Cooper S. Developing leaders for advanced life support: evaluation of a training programme. *Resuscitation* 2001;49:33–8. doi:10.1016/S0300-9572(00)00345-2

4 Malling B, Mortensen L, Bonderup T, *et al.* Combining a leadership course and multi-source feedback has no effect on leadership skills of leaders in postgraduate medical education. An intervention study with a control group. *BMC Med Educ* 2009;9:72. doi:10.1186/1472-6920-9-72

5 von Vultée PJ, Arnetz B. The impact of management programs on physicians' work environment and health: A prospective, controlled study comparing different interventions. *J Health Organ Manag* 2004;18:25–37. doi:10.1108/14777260410532047

6 Fassiotto M, Maldonado Y, Hopkins J. A long-term follow-up of a physician leadership program. *J Health Organ Manag* 2018;32:56–68. doi:10.1108/JHOM-08-2017-0208

7 Levine SA, Chao SH, Brett B, *et al.* Chief resident immersion training in the care of older adults: An innovative interspecialty education and leadership intervention. *J Am Geriatr Soc* 2008;56:1140–5. doi:10.1111/j.1532-5415.2008.01710.x

8 Hopkins J, Fassiotto M, Ku MC, *et al.* Designing a physician leadership development program based on effective models of physician education. *Health Care Manage Rev* 2018;43:293–302. doi:10.1097/HMR.0000000000000146

9 Dannels SA, Yamagata H, McDade SA, *et al.* Evaluating a leadership program: A comparative, longitudinal study to assess the impact of the Executive Leadership in Academic Medicine (ELAM) program for women. *Acad Med* 2008;83:488–95. doi:10.1097/ACM.0b013e31816be551

10 Orme D, Campbell C. How leadership training saves money 'service line leadership' at Nottingham University Hospitals. *BMJ Lead* 2019;3:29–36. doi:10.1136/leader-2018-000132

11 Cole DC, Giordano CR, Vasilopoulos T, *et al.* Resident Physicians Improve Nontechnical Skills When on Operating Room Management and Leadership Rotation. *Anesth Analg* 2017;124:300–7. doi:10.1213/ANE.0000000000001687

12 Haftel HM, Swan R, Anderson MS, *et al.* Fostering the Career Development of Future Educational Leaders: The Success of the Association of Pediatric Program Directors Leadership in Educational Academic Development Program. *J Pediatr* 2018;194:5-6.e1. doi:10.1016/j.jpeds.2017.11.066

13 Ten Have ECM, Nap RE, Tulleken JE. Quality improvement of interdisciplinary rounds by leadership training based on essential quality indicators of the Interdisciplinary Rounds Assessment Scale. *Intensive Care Med* 2013;39:1800–7. doi:10.1007/s00134-013-3002-0

14 Gilfoyle E, Gottesman R, Razack S. Development of a leadership skills workshop in paediatric advanced resuscitation. *Med Teach* 2007;29:e276–83. doi:10.1080/01421590701663287

15 LoPresti L, Ginn P, Treat R. Using a simulated practice to improve practice management learning. 2009.

16 Wurster AB, Pearson K, Sonnad SS, *et al.* The Patient Safety Leadership Academy at the University of Pennsylvania: The First Cohort's learning experience. *Qual Manag Health Care* 2007;16:166–73. doi:10.1097/01.QMH.0000267454.63123.e7

17 Pradarelli JC, Jaffe GA, Lemak CH, *et al.* A leadership development program for surgeons: First-year participant evaluation. *Surgery* 2016;160:255–63. doi:https://dx.doi.org/10.1016/j.surg.2016.03.011

18 Throgmorton C, Mitchell T, Morley T, *et al.* Evaluating a physician leadership development program - a mixed methods approach. *J Health Organ Manag* 2016;30:390–407. doi:https://dx.doi.org/10.1108/JHOM-11-2014-0187

19 Bergman D, Fransson-Sellgren S, Wahlstrom R, *et al.* Healthcare leadership: Impact of short-term intensive and long-term less intensive training programmes. *Leadersh Heal Serv* 2009;22:161–75. doi:10.1108/17511870910953805

20 Monkhouse A, Sadler L, Boyd A, *et al.* The Improving Global Health fellowship: A qualitative analysis of innovative leadership development for NHS healthcare professionals. *Global Health* 2018;14:69. doi:10.1186/s12992-018-0384-3

21 Tsoh JY, Kuo AK, Barr JW, *et al.* Developing faculty leadership from 'within': a 12-year reflection from an internal faculty leadership development program of an academic health sciences center. *Med Educ Online* 2019;24:1567239. doi:10.1080/10872981.2019.1567239

22 Bearman M, O'Brien R, Anthony A, *et al.* Learning surgical communication, leadership and teamwork through simulation. *J Surg Educ* 2012;69:201–7. doi:10.1016/j.jsurg.2011.07.014

23 Carney PA, Eiff MP, Green LA, *et al.* Transforming Primary Care Residency Training: A Collaborative Faculty Development Initiative among Family Medicine, Internal Medicine, and Pediatric Residencies. *Acad Med* 2015;90:1054–60. doi:10.1097/ACM.0000000000000701

24 Cooper JB, Singer SJ, Hayes J, *et al.* Design and evaluation of simulation scenarios for a program introducing patient safety, teamwork, safety leadership, and simulation to healthcare leaders and managers. *Simul Healthc* 2011;6:231–8. doi:10.1097/SIH.0b013e31821da9ec

25 Agius SJ, Brockbank A, Baron R, *et al.* The impact of an integrated medical leadership programme. *J Heal Organ Manag* 2015;29:39–54. doi:10.1108/JHOM-09-2013-0188

26 Mckimm J, Hickford D, Lees P, *et al.* Evaluating the impact of a national clinical leadership fellow scheme. *BMJ Lead* 2019;3:37–42. doi:10.1136/leader-2019-000135

27 Cohen D, Vlaev I, McMahon L, *et al.* The Crucible simulation: Behavioral simulation improves clinical leadership skills and understanding of complex health policy change. *Health Care Manage Rev* 2017;44:246–55. doi:10.1097/HMR.0000000000000162

28 Ruston A, Tavabie A. Fostering clinical engagement and medical leadership and aligning cultural values: An evaluation of a general practice specialty trainee integrated training placement in a primary care trust. *Qual Prim Care* 2010;18:263–8.

29 Hackworth J, Steel S, Cooksey E, *et al.* Faculty Members' Self-Awareness, Leadership Confidence, and Leadership Skills Improve after an Evidence-Based Leadership Training Program. J. Pediatr. 2018;199:4-6.e2. doi:10.1016/j.jpeds.2018.05.007

30 Al-Mutawa N, Elmahdi H, Joyce P. The implementation of a practice management programme for family medicine residents in Qatar. *Educ Prim Care* 2016;27:380–5. doi:10.1080/14739879.2016.1219620

31 Fernandez CSP, Noble CC, Jensen ET, *et al.* Improving Leadership Skills in Physicians: A 6-Month Retrospective Study. *J Leadersh Stud* 2016;9:6–19. doi:10.1002/jls.21420

32 Chang A, Lundebjerg NE, Abrams J, *et al.* Leadership, Inside and Out: The Tideswell-AGS-ADGAP Emerging Leaders in Aging Program. *J Am Geriatr Soc* 2019;67:437–42. doi:10.1111/jgs.15702

33 Day CS, Tabrizi S, Kramer J, *et al.* Effectiveness of the AAOS leadership fellows program for orthopaedic surgeons. *J Bone Jt Surg - Ser A* 2010;92:2700–8. doi:10.2106/JBJS.J.00272

34 Korschun HW, Redding D, Teal GL, *et al.* Realizing the vision of leadership development in an academic health center: The Woodruff Leadership Academy. *Acad Med* 2007;82:264–71. doi:10.1097/ACM.0b013e31803078b5

35 McDade SA, Richman RC, Jackson GB, *et al.* Effects of Participation in the Executive Leadership in Academic Medicine (ELAM) Program on Women Faculty's Perceived Leadership Capabilities. *Acad Med* 2004;79:302–9. doi:10.1097/00001888-200404000-00005

36 Edmonstone J. The development of strategic clinical leaders in the National Health Service in Scotland. *Leadersh Heal Serv* 2011;24:337–53. doi:10.1108/17511871111172376

37 MacPhail A, Young C, Ibrahim JE. Workplace-based clinical leadership training increases willingness to lead: Appraisal using multisource feedback of a clinical leadership program in regional Victoria, Australia. *Leadersh Heal Serv* 2015;28:100–18. doi:10.1108/LHS-01-2014-0002

38 Cerrone SA, Adelman P, Akbar S, *et al.* Using objective structured teaching encounters (OSTEs) to prepare chief residents to be emotionally intelligent leaders. *Med Educ Online* 2017;22:1320186. doi:10.1080/10872981.2017.1320186

39 Patel N, Brennan PJ, Metlay J, *et al.* Building the pipeline: The creation of a residency training pathway for future physician leaders in health care quality. *Acad Med* 2015;90:185–90. doi:10.1097/ACM.0000000000000546

40 Nakanjako D, Namagala E, Semeere A, *et al.* Global health leadership training in resource-limited settings: A collaborative approach by academic institutions and local health care programs in Uganda. *Hum Resour Health* 2015;13:87. doi:10.1186/s12960-015-0087-2

41 Kuo AK, Thyne SM, Chen HC, *et al.* An innovative residency program designed to develop leaders to improve the health of children. *Acad Med* 2010;85:653. doi:10.1097/ACM.0b013e3181eb60f6

42 Brandon CJ, Mullan PB. Teaching Medical Management and Operations Engineering for Systems-Based Practice to Radiology Residents. *Acad Radiol* 2013;20:345–50. doi:10.1016/j.acra.2012.09.025

43 Green PL, Plsek PE. Coaching and leadership for the diffusion of innovation in health care: A different type of multi-organization improvement collaborative. *Jt Comm J Qual Improv* 2002;28:55–71. doi:10.1016/S1070-3241(02)28006-2

44 Hemmer PR, Karon BS, Hernandez JS, *et al.* Leadership and management training for residents and fellows: A curriculum for future medical directors. *Arch Pathol Lab Med* 2007;131:610–4. doi:10.1043/1543-2165(2007)131[610:LAMTFR]2.0.CO;2

45 McCurdy FA, Beck G, Maroon A, *et al.* The administrative colloquium: Developing management and leadership skills for faculty. *Ambul Pediatr* 2004;4:124–8. doi:10.1367/1539-4409(2004)004<0124:TACDMA>2.0.CO;2

46 Hadley L, Marshall P, Black D. Pairing trainee managers and doctors: an initiative to facilitate joint working for better patient care. *Br J Hosp Med (Lond)* 2014;75:103–5.

47 Revere L, Robinson A, Schroth L, *et al.* Preparing academic medical department physicians to successfully lead. *Leadersh Heal Serv* 2015;28:317–31. doi:10.1108/LHS-03-2014-0023

48 Osborn LM, DeWitt T. The HRSA-APA Faculty Development Scholars Program: Executive Leadership Track. *Ambul Pediatr* 2004;4:98–102. doi:10.1367/1539-4409(2004)004<0098:THFDSP>2.0.CO;2

49 Wichman CL, Netzel PJ, Menaker R. Preparing psychiatric residents for the ‘real world’: A practice management curriculum. *Acad Psychiatry* 2009;33:131–4. doi:10.1176/appi.ap.33.2.131

50 Monaghan H, Swenson C, Kerins J, *et al.* Bridging the gap: using ‘Paired Learning’ to improve clinician/management understanding. *BMJ Lead* 2018;2:80–2. doi:10.1136/leader-2017-000064

51 Voogt JJ, van Rensen ELJ, van der Schaaf MF, *et al.* Building bridges: engaging medical residents in quality improvement and medical leadership. *Int J Qual Heal care J Int Soc Qual Heal Care* 2016;28:665–74. doi:https://dx.doi.org/10.1093/intqhc/mzw091

52 Heitkamp DE, Kerridge WD, Ballenger ZE, *et al.* A Leadership Development Program for Radiology Residents. *J Am Coll Radiol* 2017;14:1468–70. doi:10.1016/j.jacr.2017.05.005

53 Pearson A, Ryan C, MacVicar R. Preliminary programme evaluation of the Scottish Clinical Leadership Fellowship. *BMJ Lead* 2018;2:40–2. doi:10.1136/leader-2017-000038

54 Crites GE, Schuster RJ. A preliminary report of an educational intervention in practice management. *BMC Med Educ* 2004;4:15. doi:10.1186/1472-6920-4-15

55 Dickey C, Dismukes R, Topor D. Creating opportunities for organizational leadership (cool): Creating a culture and curriculum that fosters psychiatric leadership development and quality improvement. *Acad Psychiatry* 2014;38:383–7. doi:10.1007/s40596-014-0082-2

56 Foster T, Regan-Smith M, Murray C, *et al.* Residency education, preventive medicine, and population health care improvement: The Dartmouth-Hitchcock leadership preventive medicine approach. Acad. Med. 2008;83:390–8. doi:10.1097/ACM.0b013e3181667da9

57 Freeman AM, Nelson R, Sinha SS. The Essential Role of Leadership Development. J. Am. Coll. Cardiol. 2018;72:2272–5. doi:10.1016/j.jacc.2018.09.029

58 Saravo B, Netzel J, Kiesewetter J. The need for strong clinical leaders - Transformational and transactional leadership as a framework for resident leadership training. *PLoS One* 2017;12:e0183019. doi:https://dx.doi.org/10.1371/journal.pone.0183019

59 Schulz K, Puscas L, Tucci D. Surgical Training and Education in Promoting Professionalism: a comparative assessment of virtue-based leadership development in otolaryngology-head. *Med Educ …* 2013;18:22440. doi:https://dx.doi.org/10.3402/meo.v18i0.22440

60 Stoller JK, Berkowitz E, Bailin PL. Physician management and leadership education at the cleveland clinic foundation: Program impact and experience over 14 years. *J Med Pract Manag* 2007;22:237–42.

61 Wulfert C-H, Hoitz J, Senger U. Initial Results of the Master's Degree Programme in 'Leadership in Medicine' - Impact on hospital-based follow-on training of doctors. *GMS J Med Educ* 2017;34:Doc52. doi:https://dx.doi.org/10.3205/zma001129

62 Blumenthal DM, Bernard K, Fraser TN, *et al.* Implementing a pilot leadership course for internal medicine residents: Design considerations, participant impressions, and lessons learned. *BMC Med Educ* 2014;14:257. doi:10.1186/s12909-014-0257-2

63 Sanfey H, Harris I, Pollart S, *et al.* Evaluation of the University of Virginia Leadership in Academic Medicine Program. *Teach Learn Med* 2011;23:347–58. doi:10.1080/10401334.2011.611773

64 Edler A, Adamshick M, Fanning R, *et al.* Leadership lessons from military education for postgraduate medical curricular improvement. *Clin Teach* 2010;7:26–31. doi:10.1111/j.1743-498X.2009.00336.x

65 Richman RC, Morahan PS, Cohen DW, *et al.* Advancing women and closing the leadership gap: The executive leadership in academic medicine (ELAM) program experience. *J Women's Heal Gender-Based Med* 2001;10:271–7. doi:10.1089/152460901300140022

66 Farver CF, Smalling S, Stoller JK, *et al.* Developing leadership competencies among medical trainees: Five-year experience at the Cleveland Clinic with a chief residents' training course. *Australas Psychiatry* 2016;24:499–505. doi:10.1177/1039856216632396

67 Gregg SC, Heffernan DS, Connolly MD, *et al.* Teaching leadership in trauma resuscitation: Immediate feedback from a real-time, competency-based evaluation tool shows long-term improvement in resident performance. *J Trauma Acute Care Surg* 2016;81:729–34. doi:https://dx.doi.org/10.1097/TA.0000000000001186

68 Hill DAMD, Jimenez J-C, Cohn SM, *et al.* How To Be a Leader: A Course for Residents. *Cureus* 2018;10:e3067. doi:https://dx.doi.org/10.7759/cureus.3067

69 Pugno PA, Dornfest FD, Kahn NB, *et al.* The National Institute for Program Director Development: a school for program directors. *J Am Board Fam Pract* 2002;15:209–13.

70 Denney M, Johnstone A. Can general practice trainees engage with leadership activities during their GP training placements? - An evaluation of an intervention in South East Scotland. *Educ Prim Care* 2019;30:102–9. doi:10.1080/14739879.2019.1566782

71 McAlearney AS, Fisher D, Heiser K, *et al.* Developing effective physician leaders: changing cultures and transforming organizations. *Hosp Top* 2005;83:11–8.

72 Shah P, Cross V, Sii F. Sailing a safe ship: Improving patient safety by enhancing the leadership skills of new consultant specialist surgeons. *J Contin Educ Health Prof* 2013;33:190–200. doi:10.1002/chp.21184

73 Clapp JT, Gordon EKBB, Baranov DY, *et al.* Encouraging Reflexivity in a Residency Leadership Development Program: Expanding Outside the Competency Approach. *Acad Med* 2018;93:210–3. doi:10.1097/ACM.0000000000001915

74 Pettit JE, Dahdaleh NS, Albert GW, *et al.* Neurosurgery resident leadership development: An innovative approach. *Neurosurgery* 2011;68:546–50. doi:10.1227/NEU.0b013e318201c2ac

75 Laura Donnelly. Jeremy Hunt: Doctors and nurses should be put in charge of hospitals. Telegraph. 2016.

76 Gruver W, Spahr RC. Imparting Wisdom to Evolving Leaders. *Physician Exec* 2006;32:24–9.

77 Babitch LA. Teaching practice management skills to pediatric residents. *Clin Pediatr (Phila)* 2006;45:846–9. doi:10.1177/0009922806294216

78 Gulati K, Singh AR, Kumar S, *et al.* Impact of a leadership development programme for physicians in India. doi:10.1108/LHS-05-2019-0027

79 Stoller JK, Rose M, Lee R, *et al.* Teambuilding and leadership training in an internal medicine residency training program: Experience with a one-day retreat. *J Gen Intern Med* 2004;19:692–7. doi:10.1111/j.1525-1497.2004.30247.x

80 Edmonstone J. Evaluating clinical leadership: A case study. *Leadersh Heal Serv* 2009;22:210–24. doi:10.1108/17511870910978132

81 Stergiopoulos V, Maggi J, Sockalingam S. Teaching the physician-manager role to psychiatric residents: development and implementation of a pilot curriculum. *Acad Psychiatry* 2009;33:125–30. doi:10.1176/appi.ap.33.2.125

82 Berkenbosch L, Muijtjens AMM, Zimmermann LJI, *et al.* A pilot study of a practice management training module for medical residents. *BMC Med Educ* 2014;14:107. doi:https://dx.doi.org/10.1186/1472-6920-14-107

83 Block AA, Singh J, Kanaris AM, *et al.* Equipping our front-line managers: a national program for the Professional Development of Registrars. *Med J Aust* 2007;186. doi:10.5694/j.1326-5377.2007.tb00961.x

84 Donaghy G, McKeever K, Flanagan C, *et al.* Helping doctors in training to STEP-UP: A leadership and quality improvement programme in the belfast health and social care trust. *Ulster Med J* 2018;87:112–6.

85 Frugé E, Mahoney DH, Poplack DG, *et al.* Leadership: 'They never taught me this in medical school'. J. Pediatr. Hematol. Oncol. 2010;32:304–8. doi:10.1097/MPH.0b013e3181cf4594

86 Hunt MA, Heilman CB, Shutran M, *et al.* Commentary: An introduction to leadership self-assessment at the society of neurological surgeons post-graduate year 1 boot camp: Observations and commentary. United States: 2017. doi:10.1093/neuros/nyw093

87 Murdock J, Brammer C. A successful model of leadership development for community practice physicians. *Physician Exec* 2011;37:52-54,56.

88 Thakur A, O'Leary B, Cowie W, *et al.* The Development and Validation of a Workplace-Based Leadership Program for Senior Residents in Psychiatry. *Acad Psychiatry* 2018;43:123–7. doi:https://dx.doi.org/10.1007/s40596-018-0982-7

89 Patterson D, Godden A, Rughani A, *et al.* A leadership programme in GP training: an action learning approach. *Educ Prim Care* 2013;24:65–8. doi:10.1080/14739879.2013.11493457

90 Maza Y, Shechter E, Pur Eizenberg N, *et al.* Physician empowerment programme; a unique workshop for physician-managers of community clinics. *BMC Med Educ* 2016;16:269. doi:10.1186/s12909-016-0786-y

91 Steinert Y, Nasmith L, Daigle N. Executive skills for medical faculty: A workshop description and evaluation. *Med Teach* 2003;25:666–8. doi:10.1080/0142159032000150485

92 Satiani B, Sena J, Ruberg R, *et al.* Talent management and physician leadership training is essential for preparing tomorrow's physician leaders. *J Vasc Surg* 2014;59:542–6. doi:10.1016/j.jvs.2013.10.074

93 Lee MT, Tse AM, Naguwa GS. Building leadership skills in paediatric residents. *Med Educ* 2004;38:559–60. doi:10.1111/j.1365-2929.2004.01867.x

94 Steiner JL, Mazure C, Siggins LD, *et al.* Teaching psychiatric residents about women and leadership. 2004. doi:10.1176/appi.ap.28.3.243

95 Awad SS, Hayley B, Fagan SP, *et al.* The impact of a novel resident leadership training curriculum. *Am J Surg* 2004;188:481–4. doi:10.1016/j.amjsurg.2004.07.024

96 Bayard M, Peeples CR, Holt J, *et al.* An interactive approach to teaching practice management to family practice residents. *Fam Med* 2003;35:622–4.

97 Hanna WC, Mulder DS, Fried GM, *et al.* Training future surgeons for management roles: The resident-surgeon-manager conference. In: *Archives of Surgery*. 2012. 940–4. doi:10.1001/archsurg.2012.992

98 Patel N, Vemuri D, Frasso R, *et al.* Perceptions of Health Care Executives on Leadership Development Skills for Residents After Participating in a Longitudinal Mentorship Program. *Am J Med Qual* 2019;34:80–6. doi:10.1177/1062860618786798

99 Stefan MS, Belforti RK, Langlois G, *et al.* A Simulation-Based Program to Train Medical Residents to Lead and Perform Advanced Cardiovascular Life Support. *Hosp Pract* 2011;39:63–9. doi:10.3810/hp.2011.10.923

100 Gagliano NJ, Ferris T, Colton D, *et al.* A physician leadership development program at an academic medical center. *Qual Manag Health Care* 2010;19:231–8. doi:10.1097/QMH.0b013e3181eb13ab

101 Gurrera RJ, Dismukes R, Edwards M, *et al.* Preparing residents in training to become health-care leaders: A pilot project. *Acad Psychiatry* 2014;38:701–5. doi:10.1007/s40596-014-0162-3

102 Bhatia K, Morris CA, Wright SC, *et al.* Leadership Training for Residents: a Novel Approach. *Physician Leadersh J* 2015;2:76–80.

103 Hadley L, Black D, Welch J, *et al.* Encouraging formative assessments of leadership for foundation doctors. *Clin Teach* 2015;12:231–5. doi:https://dx.doi.org/10.1111/tct.12289

104 Kasuya R, Nip I, Anderson MB. A retreat on leadership skills for residents. *Acad Med* 2001;76:554. doi:10.1097/00001888-200105000-00099

105 Ninan D, Patel D. Career and Leadership Education in Anesthesia Residency Training. *Cureus* 2018;10:e2546. doi:https://dx.doi.org/10.7759/cureus.2546

106 Schwartz BJ, Blackmore MA, Weiss A. The Tarrytown Chief Residents Leadership Conference: A Long-Term Follow-up. *Acad Psychiatry* 2014;38:15–8. doi:10.1007/s40596-013-0016-4

107 Ennis-Cole DL, Cullum PM, Iwundu C. Physicians as Operational Leaders: Cost, Curriculum, Technology, and Organizational Challenges. *TechTrends* 2018;62:239–49. doi:10.1007/s11528-018-0273-x

108 Steinhardt L. Workshop for New Leaders: Innovative Midwifery Teaching for Obstetrics and Gynecology Residents. *J Midwifery Women's Heal* 2015;60:313–7. doi:10.1111/jmwh.12292

109 Torbeck L, Rozycki G, Dunnington G. Leaders Growing Leaders: Designing a Tier-Based Leadership Program for Surgeons. *J Surg Educ* Published Online First: 2018. doi:https://dx.doi.org/10.1016/j.jsurg.2017.12.009

110 Vimr M, Dickens P. Building Physician Capacity for Transformational Leadership—Revisited. *Healthc Manag Forum* 2013;26:16–9. doi:https://doi.org/10.1016/j.hcmf.2013.01.003

111 Bircher J. Extending GP training and the development of leadership skills: The experience of the North Western Deanery pilot. *Educ Prim Care* 2013;24:57–60. doi:10.1080/14739879.2013.11493457

112 Kochar MS, Robertson RG, Mone MA. A faculty leadership development program at the Medical College of Wisconsin. Wis. Med.

J. 2003;102:24–8.

113 Biese K, Leacock BW, Osmond CR, *et al.* Engaging Senior Residents as Leaders: A Novel Structure for Multiple Chief Roles. *J Grad Med Educ* 2011;3:236–8. doi:10.4300/jgme-d-10-00045.1

114 Cherry RA, Davis DC, Thorndyke L. Transforming culture through physician leadership development. *Physician Exec* 2010;36:38–44.

115 Johnson JM, Stern TA. Teaching Residents About Emotional Intelligence and Its Impact on Leadership. *Acad Psychiatry* 2014;38:510–3. doi:10.1007/s40596-014-0048-4

116 O'donnell L, Bennett GL. Case management takes lead role in educating medical residents. *Prof Case Manag* 2011;16:253–5. doi:10.1097/ncm.0b013e31821b0785

117 Ringdahl EN, Tarwater KD, Lindbloom EJ. A Longitudinal Curriculum to Address the Gender Gap in Physician Leadership. *J Grad Med Educ* 2014;6:361–2. doi:10.4300/JGME-D-14-00081.1

MERSQI Component	Classification	All studies (117)	MERSQI>12 (16)	JBI>6 (14)
Study Design	Single Group Cross-Sectional or Post-programme only	54 (46%)	0 (0%)	8 (57%)
	Single Group Pre and Post Programme	54 (46%)	9 (56%)	5 (36%)
	Non-Randomised Two Group	8 (7%)	6 (38%)	0 (0%)
	Randomised Controlled Trial	1 (1%)	1 (6%)	1 (7%)
Institution #	Single	81 (69%)	10 (63%)	9 (64%)
	Double	1 (1%)	1 (6%)	0 (0%)
	Multi	35 (30%)	5 (31%)	5 (36%)
Response Rate	<50% or NR	44 (38%)	3 (19%)	1 (7%)
	50-75%	23 (20%)	3 (19%)	6 (43%)
	>75%	48 (41%)	9 (56%)	7 (50%)
Type of Data	Self-reported	70 (60%)	2 (13%)	9 (64%)
	Observed	47 (40%)	14 (88%)	5 (36%)
Questionnaire Construct Validity (Internal Structure)	Reported	9 (8%)	7 (44%)	2 (14%)
	Not Reported	108 (92%)	9 (56%)	12 (86%)
Questionnaire Content Validity (Content)	Reported	45 (38%)	14 (88%)	5 (36%)
	Not Reported	72 (62%)	2 (13%)	9 (64%)
Relationships to Other Variables	Reported	8 (7%)	5 (31%)	3 (21%)
	Not Reported	108 (92%)	11 (69%)	11 (79%)
Data Analysis Comprehensiveness (Appropriateness)	Comprehensive	23 (20%)	14 (88%)	6 (43%)
	Less Comprehensive	94 (80%)	2 (13%)	8 (57%)
Data Analysis Complexity	Descriptive only	102 (87%)	7 (44%)	11 (79%)
	Beyond Descriptive	15 (13%)	9 (56%)	3 (21%)
Outcomes (Kirkpatrick Level)	Level 1	80 (68%)	8 (50%)	14 (100%)
	Level 2a	70 (60%)	7 (44%)	13 (93%)
	Level 2b	79 (68%)	11 (69%)	11 (79%)
	Level 3a	51 (44%)	7 (44%)	10 (71%)
	Level 3b	54 (46%)	14 (88%)	7 (50%)
	Level 4a	9 (8%)	1 (6%)	2 (14%)
	Level 4b	26 (22%)	7 (44%)	4 (29%)

Table 18: Study characteristics organised by Medical Education Research Study Quality Indicator (MERSQI) heading.
Brackets in headings refer to original MERQSI items where headings have been adapted for clarity.

JBI Component	Description of component	High-reliability studies (n=14)	Mixed-Methods Studies (n=53)	Qualitative Studies (n=10)
Philosophical Perspective	*Congruity between the stated philosophical perspective and the research methodology*	0 (0%)	0 (0%)	0 (0%)
Objectives	*Congruity between the research methodology and the research question or objectives*	39 (63%)	33 (62%)	6 (60%)
Methods	*Congruity between the research methodology and the methods used to collect data*	38 (61%)	32 (60%)	6 (60%)
Analysis	*Congruity between the research methodology and the representation and analysis of data*	18 (29%)	15 (28%)	3 (30%)
Interpretation	*Congruity between the research methodology and the interpretation of results*	17 (27%)	13 (25%)	4 (40%)
Researcher Location	*Statement locating the researcher culturally or theoretically*	10 (16%)	8 (15%)	2 (20%)
Researcher Influence	*Influence of the researcher on the research addressed*	13 (21%)	12 (23%)	1 (10%)
Participant Representation	*Participants and their voices adequately represented*	21 (34%)	16 (30%)	5 (50%)
Ethics	*Evidence of ethical approval by an appropriate body*	26 (42%)	23 (43%)	3 (30%)
Conclusions Supported	*Conclusions drawn in the research report flow from the analysis or interpretation of the data*	20 (32%)	16 (30%)	4 (40%)

Table 19: Proportion of studies which met Joanna Briggs Institute (JBI) Critical Appraisal Items for Qualitative Studies. Descriptions are adapted from the JBI tool. Higher Reliability Studies scored 6 or more on the JBI tool

# ▲	Searches	Results	Type	Actions	Annotations
1	PHYSICIANS/	87338	Advanced	Display Results | More	
2	PHYSICIAN EXECUTIVES/	4211	Advanced	Display Results | More	
3	CONSULTANTS/	6648	Advanced	Display Results | More	
4	INTERNSHIP/	48084	Advanced	Display Results | More	
5	RESIDENCY/	48084	Advanced	Display Results | More	
6	MEDICAL STAFF/	2559	Advanced	Display Results | More	
7	(physician* or surgeon* or doctor* or intern or interns or residen* or registrar* or consultant* or "house officer*" or "medical staff*").ti.	230256	Advanced	Display Results | More	
8	1 or 2 or 3 or 4 or 5 or 6 or 7	304607	Advanced	Display Results | More	
9	LEADERSHIP/	40117	Advanced	Display Results | More	
10	PRACTICE MANAGEMENT/	1344	Advanced	Display Results | More	
11	"leader*".ti.	21361	Advanced	Display Results | More	
12	(practice and manag*).ti.	9878	Advanced	Display Results | More	
13	9 or 10 or 11 or 12	58933	Advanced	Display Results | More	
14	PROGRAM EVALUATION/	61451	Advanced	Display Results | More	
15	PROGRAM DEVELOPMENT/	28428	Advanced	Display Results | More	
16	CURRICULUM/	73527	Advanced	Display Results | More	
17	EDUCATION,MEDICAL,CONTINUING/	24544	Advanced	Display Results | More	
18	EDUCATION,MEDICAL,GRADUATE/	28493	Advanced	Display Results | More	
19	(teach* or train* or educat* or course* or program* or pathway* or curricul*).ti.	783506	Advanced	Display Results | More	
20	14 or 15 or 16 or 17 or 18 or 19	873227	Advanced	Display Results | More	
21	8 and 13 and 20	1068	Advanced	Display Results | More	
22	limit 21 to yr="2013 -Current"	461	Advanced	Display Results | More	
23	GENERAL PRACTITIONERS/	7485	Advanced	Display Results | More	
24	FAMILY PHYSICIANS/	16222	Advanced	Display Results | More	
25	MEDICAL STAFF,HOSPITAL/	22748	Advanced	Display Results | More	
26	(GP or GPs or "general practitioner*" or "family practitioner*").ti.	22053	Advanced	Display Results | More	
27	23 or 24 or 25 or 26	60255	Advanced	Display Results | More	
28	13 and 20 and 27	161	Advanced	Display Results | More	
29	limit 28 to yr="2013 -Current"	63	Advanced	Display Results | More	
30	29 not 22	40	Advanced	Display Results | More	

Figure 14: Medline (OVID) Search Strategy for systematic review in Chapter Ten, January 2020

Appendix Three – Original Primary Colours Questionnaire

Oxford University Hospitals Future Leaders Programme. Post programme survey July 2017

A leadership programme centred around a quality improvement project is necessarily complex and this questionnaire is designed to gather as much information as we can about the Future Leaders Programme. This means quite a few questions! As principal members of the project, your role in evaluation is crucial. Understanding how to make future programmes more effective is our principal aim and is definitely **not** identifying individual or group success or lack of success. **Your responses are anonymised** as they were for the evaluation we did in January. ***FLP*** refers to Future Leaders Programme.
Some of the questions are about your **participation as leaders** in your own project, and others are about **faculty effectiveness** and **programme design.** So the questionnaire has three strands. **1. Your experience, effectiveness and learning. 2. Faculty programme and effectiveness. 3. Effectiveness of project direction and oversight.** You will notice these three types of questions as you complete the questionnaire. We will sort them out in the analysis. The questionnaire is similar to the one used for last year's **FLP.** Some questions do not apply to your programme and have been struck through.
COMPLETING THE QUESTIONNAIRE: This is a "forced choice three option" format to try and obtain clearer results so in choosing your answer, try and select the answer that is the closest to your view. This hopefully captures an element of decision making on your part that would otherwise be lost. ***If you cannot decide between two options, then please ring two e.g.: ring "unsure" AND "yes" to indicate your answer lies between the two. There is space for you to add in brief comments but don't think you have to add any. Your answers to the question is what we are principally interested in obtaining. Thank you so much.***

STRATEGIC DOMAIN ***Definition: "a plan designed to achieve a particular long-term aim"***				**COMMENT**
Q1. Did **you** have any understanding of the organisations strategic aims when you first started the Future Leaders Programme last October?	NO	UNSURE	YES	
Q2 Do **you** now have an understanding of the organisation (OUH) strategic aims?	NO	UNSURE	YES	
Q3 Was the **executive team** clear in articulating the organisations strategic aims?	NO	UNSURE	YES	
Q4 Was the **programme and/or faculty** helpful in articulating the importance of understanding the organisations strategic aims?	NO	UNSURE	YES	
Q5 Were **you** able to articulate and employ the organisations	NO	UNSURE	YES	

strategic aims to help you develop your own project?

STRATEGIC ALIGNMENT

Definition: "the process of building commitment to the organisations aims is what is called 'creating alignment'."

Q6. Do **you** believe that your **FLP project** was aligned to the organisations strategic aims?	NO	UNSURE	YES
Q7. Do **you** believe your **project sponsors** understood your *FLP project* was aligned to the organisations strategic aims?	NO	UNSURE	YES
Q8. Do **you** believe your **directorate** understood your *FLP project* was aligned to the organisations strategic aims?	NO	UNSURE	YES
Q9. Do you think the **programme and faculty** gave you the knowledge and tools to understand the importance of and the means to utilise alignment to help you in your project?	NO	UNSURE	YES

OPERATIONAL DOMAIN

Definition: "delivering results represents the ability to drive work through to completion"

Q10. How supported were **you** by the **Executive Team** to drive the *FLP project* work through to completion?	NO SUPPORT	UNSURE	YES GOOD SUPPORT
Q11.How supported were **you** by the **Project sponsor** to drive the *FLP project* through to completion?	NO SUPPORT	UNSURE	YES GOOD SUPPORT
Q12. How supported were **you** by **your directorate** to drive the *FLP project* through to completion?	NO SUPPORT	UNSURE	YES GOOD SUPPORT
Q13. How supported were you by the **Faculty and programme** to drive the FLP project through to completion?	NO SUPPORT	UNSURE	YES GOOD SUPPORT
Q14. How motivated did **you** feel to drive the *FLP project* through to completion?	NO MOTIVATION	UNSURE	WELL MOTIVATED
Q15. How motivated do you think **your FLP tea*m*** was to drive the *FLP project* through to completion?	NO MOTIVATION	UNSURE	WELL MOTIVATED
Q16. How responsible did **you** personally feel to drive the *FLP project* through to completion?	NOT VERY RESPONSIBLE	UNSURE	VERY RESPONSIBLE

PLANNING AND ORGANISATION

Definition: "an organisation achieves results with plans broken down into objectives and tasks"

Q17. Do **you** think that your **FLP project** was appropriately broken down into objectives and tasks?	NO	UNSURE	YES
Q18. Do **you** think the **faculty and programme** provided you with the knowledge and tools to produce a credible plan to implement your **FLP** project?	NO	UNSURE	YES
Q19. Do **you** think appropriate time and effort was spent by **your team** on planning and organising for the **FLP projec**t?	NO	UNSURE	YES
Q20. Do you think appropriate time and effort was spent by **yourself** on planning and organising for your **FLP project**?	NO	UNSURE	YES
Q21. Do you think the quality of input and interaction with **the project sponsor and the steering committee** was sufficient to enable you to plan and organise your **FLP project**?	NO	UNSURE	YES
Q22. Do you think the quality of input and interaction with the **faculty and programme** was sufficient to enable you to plan and organise your **FLP project**?	NO	UNSURE	YES

INTERPERSONAL DOMAIN

Definition: "building and sustaining relationships is the core activity of the interpersonal domain"

Q23. Did **you** feel that you successfully built and maintained relationships with the **team members** of your own **FLP project**?	NO	UNSURE	YES
Q24. Do **you** think the **allocation of team members** to individual projects was appropriate?	NO	UNSURE	YES
Q25. Did you feel that **you** successfully built and maintained relationships with your **project sponsor**?	NO	UNSURE	YES
Q26. Did you feel **you** successfully built and maintained relationships with your **main FLP projec*t*** stakeholders?	NO	UNSURE	YES
Q27. ~~Did you feel **you** successfully built and maintained relationships with your **strategy project** stakeholders?~~	NO	UNSURE	YES

TEAM BUILDING

Definition: "a set of values, standards and behaviours that hold a group engaged on a task together"

Q28. Did the **faculty and programme** provide you with the knowledge and skills to help develop team building?	NO	UNSURE	YES
Q29. How well did **you** think your **FLP project team** functioned?	POORLY	UNSURE	WELL
Q30. ~~How well did you think your **strategy project** team functioned?~~	POORLY	UNSURE	WELL
Q31. Was sufficient time and opportunity given to **you** for your **FLP project** team to meet?	NO	UNSURE	YES
Q32. ~~Was sufficient time and opportunity given to **you** for your **strategy project** team to meet?~~	NO	UNSURE	YES
Q33. How much extra effort did **you** have to make over and above your extra PA to make your **FLP** progress?	NONE	UNSURE	Greater than 1PA

INSPIRATION

Definition: "the means by which others become motivated to buy in to making change"

Q34. How confident were you in the **competence** of the **faculty and programme** to deliver knowledge and tools needed to result in change?	NOT AT ALL	UNSURE	CONFIDENT
Q35. Did you feel the **faculty and programme** really cared about the success of the project?	NO	UNSURE	YES
Q36. Did the **faculty and programme** have integrity in terms of their consistent behaviour and values towards you?	NO	UNSURE	YES
Q37. Did **you** believe that the **faculty and programme** was honest with you about your own progress in your **FLP projects**?	NO	UNSURE	YES
Q38. Did **you** believe that your **project sponsor** was honest with you about your own progress in your **FLP project**?	NO	UNSURE	YES
Q39. Did **you** feel engaged with the **faculty and programme** in your efforts to complete the **FLP projects**?	NO	UNSURE	YES
Q40. Did **you** feel engaged with the **project sponsor** in your efforts to complete the **FLP project**?	NO	UNSURE	YES
Q41. Did **you** feel engaged with **your directorate** in your efforts to complete the **FLP project**?	NO	UNSURE	YES
Q42. Did **you** think you have acquired some skills to inspire	NO	UNSURE	YES

others to engage with you in your **own project for example, with Foundation trainees**?

FOCUS

Definition: "the ability to know where to place effort for maximum benefit towards completion of the task"

Q43. Do **you** think the **faculty and programme** have **now** provided you with the knowledge and skills to develop objectives, priorities, plans and schedules in order to complete your project?	NO	UNSURE	YES
Q44. Did **you** think the **faculty and programme** paid critical attention to objectives, priorities, plans and schedules in order to help complete your **FLP projects**?	NO	UNSURE	YES
Q45. Did **you** think **project sponsor** paid critical attention to objectives, priorities, plans and schedules in order to complete your project?	NO	UNSURE	YES
Q46. Did **you** think that the e**xecutive Team** was clear about what needed to be done and timeframes needed to complete the tasks?	NO	UNSURE	YES
Q47. Did **you** think that the **project sponsor** was clear about what needed to be done and timeframes needed to complete the tasks?	NO	UNSURE	YES
Q48. Did **you** think that the **faculty and programme** was clear about what needed to be done and timeframes needed to complete the tasks?	NO	UNSURE	YES
Q49. Did **you** feel that **external factors** (the urgent, emergencies etc.) significantly interfered in your ability to complete your task?	YES	UNSURE	NO
Q50. Did **you** feel that **bureaucracy, repetition, or routine** played a role in preventing you from completing your tasks?	YES	UNSURE	NO
Q51. Did **you** feel that the **difficulties in meeting together** played a role in preventing you from completing your tasks?	YES	UNSURE	NO
Q52. How well do **you** think you executed your task in terms of focusing on objectives, priorities, plans and schedules in your own **FLP** project?	POORLY	UNSURE	WELL

ENABLE

Definition: "the ability to give others the mandate, resources and skills needed to act"

Q53. How empowered did **you** feel to be able to continue and complete the **FLP** project?	NOT EMPOWERED	UNSURE	EMPOWERED
Q54. ~~How empowered did **you** feel to be able to complete the **strategy project**?~~	NOT EMPOWERED	UNSURE	EMPOWERED
Q55. Did **you** ever feel able to act in a "self-enabled" manner by doing what you considered right for the organisation and take the consequences should there be any?	NOT SELF ENABLED	UNSURE	SELF ENABLED
Q56. Did **you** ever consider or try to enable your team in your **own project for example, with Foundation trainees?**	NO	UNSURE	YES

REINFORCE

Definition: "a process of performance management that ensures that tasks are completed"

Q57. Did **you** feel committed to ensuring that the tasks were completed for the **FLP** project?	NO	UNSURE	YES
Q58. ~~Did **you** feel committed to ensuring that the tasks were completed for the **strategy** project?~~	NO	UNSURE	YES
Q59. Did **you** feel that the **project sponsor** supported you when the project encountered lack of progress?	NO	UNSURE	YES
Q60. Did **you** feel that your **directorate** supported you when the project encountered lack of progress?	NO	UNSURE	YES
Q61. Did **you** feel that the **faculty and programme** supported you when the project encountered lack of progress?	NO	UNSURE	YES
Q62. Did **you** feel that the level of support generally was sufficient?	NO	UNSURE	YES

Learn

Definition: "an organisation that wants to learn and its people to learn"

Q63. Do **you** think **faculty and programme** gave you sufficient opportunities to learn and reflect on what was	NO	UNSURE	YES

happening?

Q64. Do **you** think ***your project sponsor*** gave you sufficient opportunities to learn and reflect on what was happening?	NO	UNSURE	YES
Q65. Do **you** think that the skills **you learn**t was worth being part of the programme?	NO	UNSURE	YES
Q66. Do **you** think the skills you learnt **was worth the effort** you put into attending the programme?	NO	UNSURE	YES
Q67. Are **you** glad you joined the programme?	NO	UNSURE	YES
Q68. Would **you** recommend the programme to a colleague?	NO	UNSURE	YES

Please feel free to express your personal opinion on the FLP project.
Could you write a few words on the positive outcomes that are a consequence of taking part in the programme?

Could you write a few words on the negative outcomes that are a consequence of taking part in the programme?

THANK YOU.
Oscar Lyons and Richard Canter
Saltford Manor,
Queens Square
Bristol. BS31 3EL
Richard.canter@nds.ox.ac.uk
O7718 989371

Appendix Four – Revised Primary Colours Questionnaire

Emerging Leaders Programme Evaluation July 2018

Explanation:
Our principal aim is to understand how to make both this programme and future leadership programmes more effective. In order to do this, we invite you to be co-researchers in this process, and we will constantly seek your advice, ideas, and suggestions.

The data collected will increase our understanding of how to improve management and leadership programmes. This particular questionnaire is also part of the evaluation we are expected to do for Oxford University Hospitals because the programme involves the use of public funds.

All questionnaires will be anonymized. Only you will have access to your own code when collated data is viewed by yourself and others.

All data collected will be stored securely. This will include your questionnaires, presentations and feedback.

You will be free to withdraw from this project at any time should you wish to do so.

If you have any questions or concerns, please contact Oscar using the contact details below.

Instructions
Please circle the number which corresponds to your rating (1=very poor, 10=excellent)

Very poor Adequate Excellent

1 2 3 4 5 6 7 8 9 10

Space has been left at the end for further comments, which you should feel free to make wherever appropriate.

Thank you very much for your input,

Oscar Lyons and Richard Canter
Balliol College
OX13BJ
Oxford
oscar.lyons@nds.ox.ac.uk
07491868963

Participant Code:

Domains of leadership

Setting Strategic Direction

How would you rate your knowledge of trust strategy?

1 2 3 4 5 6 7 8 9 10

Creating Alignment

How would you rate your commitment to the trust's aims?

1 2 3 4 5 6 7 8 9 10

Building and sustaining relationships

How would you rate your relationship with hospital managers? (for example, clinical directors, divisional directors, medical director)

1 2 3 4 5 6 7 8 9 10

How would you rate your team working relationships with your peers?

1 2 3 4 5 6 7 8 9 10

How would you rate your relationship with nursing staff?

1 2 3 4 5 6 7 8 9 10

How would you rate your relationship with other staff?

1 2 3 4 5 6 7 8 9 10

Team working

How would you rate your ability to increase discretionary effort in a team?

1 2 3 4 5 6 7 8 9 10

Planning and organising

How would you rate your ability to break down work into objectives and tasks?

1 2 3 4 5 6 7 8 9 10

Delivering Results

How would you rate your ability to generate valid and measurable outputs for projects?

1 2 3 4 5 6 7 8 9 10

Tasks of leadership
Inspire: motivating people to buy in to change
How would you rate your ability to inspire others?

1 2 3 4 5 6 7 8 9 10

Focus: knowing where to place effort for maximum benefit
How would you rate your ability to focus others?

1 2 3 4 5 6 7 8 9 10

Enable: equipping people to handle the discretion they are given and supporting their actions.
How would you rate your ability to enable others?

1 2 3 4 5 6 7 8 9 10

Reinforce: ensuring that good performance is rewarded, those who are struggling are helped, and persistent poor performance is appropriately addressed
How would you rate your ability to reinforce performance in others?

1 2 3 4 5 6 7 8 9 10

Learn: showing commitment to learning and development, and fostering a learning environment
How would you rate your ability to help others learn?

1 2 3 4 5 6 7 8 9 10

Further questions
What factors have particularly affected your ability to implement change (positive and negative)

What is your gender?

Further comments:
Please add any further comments here (positive, negative and otherwise)
for example, areas you would like to focus on/feel are important, ideas for improvement.

Appendix Five – Supplementary Questionnaire

Oxford University Hospitals Future Leaders Programme survey September 2017

Participant Code:
Explanation:
We would like to gain as full an insight as possible to inform development of the FLP (Future Leaders Programme). We have had feedback from participants indicating there were outcomes which we may have overlooked. With participant input we have developed this short questionnaire to supplement previous surveys.
Our intention is to capture both informal and more formalised use of what you learnt on the programme

Completing the questionnaire*:*
As in the previous survey, there are a small number of 3-option questions, designed to force a choice. Space has been left for comments in certain questions, but please feel free to add comments wherever appropriate.

Survey

Skill application outside FLP Project

1. During the FLP, have you found yourself applying skills learned in the Programme in areas outside your main project? (for example, in clinical activities, departmental activities, in your personal lives, charity work or any other)	**No**	**Unsure**	**Yes**

Circle as many as appropriate:
Clinical activity - Departmental – Personal lives – Charity work - Other (please specify)

2. What have you learnt that you have found particularly useful? (for example, stakeholder analysis, negotiation and influencing skills, personality profiling, self-reflection, developing a business case/pitch, presentation skills, team working etc)
Comment (learning/skill and where you used them):

Clinical improvement projects outside FLP

3. Since you started the FLP, have you been involved or started clinical improvement projects other than your FLP project?	**No**	**Unsure**	**Yes**
4. Have you found yourself spending more time on the FLP project or on other projects?	**FLP**	**Unsure or n/a**	**Other**
5. Has your learning from the FLP been more useful on the FLP project or on other projects?	**FLP**	**Unsure or n/a**	**Other**
6. Which project has in your opinion had a greater impact? *Comment:*	**FLP**	**Unsure or n/a**	**Other**

Purpose and future of FLP

7. What in your opinion seemed to be the main goal of the FLP?
Comment:

8. In your opinion, has this goal been achieved? *Comment:*	**No**	**Unsure**	**Yes**

9.	Do you expect your FLP project to continue after the official end of the programme?	**No**	**Unsure**	**Yes**

Personal impact of FLP

10.	Has the FLP affected your confidence (for example, personal/team/departmental interactions, with more senior management)	**Decreased**	**Unsure**	**Increased**
11.	Has the FLP affected your motivation (for example, personally, within your work/team/department)	**Decreased**	**Unsure**	**Increased**
12.	Has the FLP affected your job satisfaction	**Decreased**	**Unsure**	**Increased**
13.	Has the FLP affected your resilience	**Decreased**	**Unsure**	**Increased**

Use of information in evaluation, development and research

This questionnaire is part of the evaluation we are expected to do for OUH. We are anticipating that the data collected today will form part of further research to improve management and leadership programmes.
Should this project form part of a research programme, ethical approval will be obtained, and we will contact you to complete any further documentation required. You will be free to withdraw from such a project at any time should you wish to do so.
Specific data collected will include your questionnaires, presentations and feedback, which will be stored securely.
All questionnaires have been anonymized.

Further comments

If you have any further comments, particularly regarding projects outside the FLP and also any other comments, include these here
If you would like to have a 30min conversation to elaborate, please email Oscar Lyons at oscar.lyons@nds.ox.ac.uk
Comment:

Thank you very much for your input

Oscar Lyons and Richard Canter
Saltford Manor,
Queens Square
Bristol. BS31 3EL
Richard.canter@nds.ox.ac.uk

Appendix Six – Case study descriptions

Programme Name	Abbreviation	Learners	Length	Location	Main Evaluation Methods*
Future Leaders Programme (2016-2017)	*FLP 2016-2017*	24 newly qualified medical consultants (n=24)	*12 months*	*On-site (Satellite building)*	*Original Primary Colours Questionnaire* *Supplementary Questionnaire*
Emerging Leaders Programme (Cohort 1)	*ELP1*	14 junior doctors (F1-CT2) on representative committees	*5 months*	*On-site (Main building education centre)*	*Revised Primary Colours Questionnaire* *Supplementary Questionnaire*
Emerging Leaders Programme (Cohort 2)	*ELP2*	11 medical trainees 2 nurses 2 midwives 2 physiotherapists	*8 months*	*On-site (Main building education centre)*	*Revised Primary Colours Questionnaire* *MLCF Self-Assessment* *Supplementary Questionnaire*
Emerging Leaders Programme (Cohort 3)	*ELP3*	13 medical trainees 4 nurses 1 occupational therapist 1 physiotherapist 1 research manager	*8 months*	*Off-site (New College Oxford Conference Room)*	*Revised Primary Colours Questionnaire* *MLCF Self-Assessment* *Supplementary Questionnaire* *Brief Resilience Scale*
IMPACT - Improving Performance and Care in Teams	*IMPACT*	8 Senior management triumvirates (each with a Matron, Clinical Director and Operational Service Manager)	*12 months*	*Off-site (Kassam Stadium Conference Centre)*	*Interviews*

Table 20: Case study details

All case studies were developed with experiential learning as a focus. They had a mixture of workshops and longitudinal team leadership projects focused on quality improvement.

Workshops for all case studies included sessions with inspirational guest speakers from inside and outside the organisation, sessions focused on leadership skills and knowledge, and sessions for facilitated group project work.

All workshops were full-day workshops except ELP1, which had half-day workshops. Topics for workshop sessions varied slightly between programmes but were focused on practical leadership skills, within a distributed leadership model (Primary Colours).

The Future Leaders Programme and Emerging Leaders Programmes had monthly workshops; the IMPACT programme had one workshop every two months.

Emerging Leaders Programme Cohort Three Overview

As an example of the design of the case studies, Emerging Leaders Programme (Cohort Three/ELP3) is outlined below. Each of the case studies (FLP, ELP and IMPACT) varied slightly cohort-cohort and programme-programme according to the needs of participants, feedback from participants and the availability of speakers. The general structure and content was similar to the programme described below.

Programme Overview
Over 7 months, ELP3 participants attended monthly sessions with invited speakers, lectures, workshops and skills sessions. They explored leadership, management and sustainable change, and engage in discussions with senior leaders who offered insights from their personal experiences. Throughout the programme participants worked together in teams on a range of quality improvement projects to put into practise knowledge and skills learned on the programme.

Programme Participants
ELP3 invited multidisciplinary applicants from all healthcare professions, healthcare management, and both primary and secondary care.

Overall Aim
The aim of ELP3 was to intervene at a critical career stage to provide a strong, general foundation in leadership for participants to build on as their careers progress.

Participants

Number:
Up to 24 multidisciplinary participants

Description:
Healthcare professionals working in the Oxford University Hospitals Foundation Trust
Doctors from Foundation, Internal Medicine and Specialty Training representative groups, from Academic Foundation Training and Oxford Deanery Quality Improvement Fellowships.
Nursing and Allied Health clinicians at Band 6 (team leader) and above.
Applications were also accepted from Quality Improvement hub programme graduates.

Cost to participants
Successful applicants were allocated a place on the programme free of charge.

Recruitment
Applications for the programme were invited via an online form. This application included:

1. Programme information:
    a. Programme description
    b. Participant description
2. Application Information
    a. Availability for dates of programme workshops
    b. Reason for interest in the programme
    c. Past leadership development experience

Programme and Curriculum

This programme used blended learning techniques to introduce and develop leadership and teamwork skills. Teaching methods included presentations, small and large group discussions and facilitated project work. Participants were encouraged to actively employ learning in their environments through team quality improvement projects, affording them an opportunity to experiment with and adapt new skills and knowledge.

Structure:

- Monthly workshops (seven full days)
- Team projects (topics to be decided within teams, in consultation with Oxford University Hospitals Stakeholders)

Workshops:
To allow time for team projects, there were seven full-day sessions spaced over eight months.
Applicants agreed to attend at least six of seven sessions including the first and last session and had approval to do so.

Project work:
Team leadership projects were centred on quality improvement, and offered an opportunity for participants to work together as a leadership team, to apply their learning from the programme, and to develop as collaborative leaders.
Teams were expected to work on their projects between sessions, and to meet weekly for ~30min to progress these.
These team meetings could be in person, via video calling or phone calls, to enable team members to meet flexibly and plan their project progress.
Team projects were presented to the trust during the final workshop and at the hospital's quality improvement symposium.

Workshop Structure

Full day workshops:
0900-0915 Tea and coffee
0915-1000 Invited Speaker (internal or external)
1000-1045 Teaching Session One
1045-1100 Break
1100-1230 Teaching Session Two
1230-1315 Lunch
1315-1400 Teaching Session Three
1400-1600 Project Group Work

Workshop Topics

Topics were adjusted according to participant needs

Session One
1. Programme Outline and Introductions
2. Healthcare Leadership
3. Building Better Teams: The fundamentals of team dynamics
4. Personality and working styles
5. Stakeholder analysis (influence mapping)
6. Project management

Tasks before next session: contact key stakeholders, plan short baseline evaluation

Session Two
1. Understanding and acting on patient experiences
2. Building sustainable change
3. Process mapping, analysing clinical care pathways
4. Sources of insight

Tasks before next session: contact source of insight, complete baseline evaluation

Session Three
1. Healthcare systems – politics and finances
2. Evaluating impact

Tasks before next session: achieve "quick win"

Session Four
1. Understanding conflict styles
2. Mediation skills: recognising, addressing and deescalating conflict
3. Negotiation

Tasks before next session: apply an influence or negotiation skill to the team project

Session Five
1. Business plans and health economics
2. Process-based decision making
3. Influence principles and tactics

Tasks before next session: draft business plan, apply influence tactic and/or decision-making tool.

Session Six
1. Effective Presentations
2. Media Interviews

Tasks before next session: prepare poster and presentation

Session Seven
1. Excellence
2. Appreciative management
3. Project Presentations (Afternoon)

QI Symposium
1. Project poster presentations