NETWORKS FOR SCHOOL SUPPORT AND IMPROVEMENT:

A mixed methods study

Ariel Mariah Lindorff

Department of Education

University of Oxford

Harris Manchester College

Supervisor: Professor Pam Sammons

Thesis submitted to the University of Oxford in fulfilment of the requirements for the degree of Doctor of Philosophy in Education

Trinity Term, 2016

ABSTRACT

From 2010 to 2016, public schools in New York City were required to affiliate with one of approximately 60 Children First Networks (CFN). These networks were designed to provide a broad range of support for instruction, leadership, curriculum, and operations in schools, as well as facilitating collaboration between schools, with the ultimate goal of raising student achievement. Using mixed methods, the research presented in this thesis investigates relationships between these networks and student achievement, and explores processes and interactions relevant to collaboration and school improvement within and between schools and their affiliated networks.

The overall explanatory sequential mixed methods design began with a quantitative phase, in which secondary data were analyzed using multilevel modeling to determine whether there was an overall effect of networks on student achievement on state assessments in literacy and mathematics. Contextualised value added (CVA) school effects were also investigated at this stage, and the results were used to inform the sampling of schools for the qualitative strand of the project. Qualitative inquiry used a case study approach, with the "case" defined as the network structure as a whole, and embedded cases consisting of 10 schools in several different networks. Interviews were conducted with various stakeholders (principals, teachers, and network staff) to access multiple perspectives and experiences of network roles, participation and collaboration in networks, and how these relate to school improvement.

Main findings from the quantitative analysis demonstrate little evidence of an overall effect of networks on student achievement, but show significant variation between schools' CVA effects. Qualitative findings offer insight relevant to these results, as accounts of participants at different levels of the network structure vary regarding the nature, extent and impact of schools' involvement with networks.

Focusing on an under-researched network structure and applying rigorous methods from educational effectiveness research within a mixed methods design to study networking and collaboration in education, this study makes an original and substantive contribution to both of these fields. Findings have the potential to inform relevant policy in New York City, and though results from this context must be generalized cautiously, this research may provide useful considerations for policymakers in other settings by adding to the broader evidence base on the effects of school networks and the processes and interactions within them.

Acknowledgements

I would like to express deep thanks to my supervisor, Professor Pam Sammons, whose expertise, encouragement and guidance were invaluable. I feel fortunate and deeply grateful to have had such excellent and also caring supervision over the course of this doctoral journey.

I would also like to thank Professor Ian Menter, who was my second supervisor and whose advice and guidance was especially helpful in the early stages of this research with regard to the structure of the thesis and my engagement with theory.

To the family and friends who were supportive throughout the many highs and lows of this process, I am thankful to have had a wonderful, caring, kind and patient support system to keep me going during the more taxing moments of this process.

TABLE OF CONTENTS

ABSTRACT I
TABLE OF CONTENTS III
LIST OF TABLES VI
LIST OF EQUATIONS IX
LIST OF FIGURES X
LIST OF ABBREVIATIONS XII

1 INTRODUCTION 1
1.1 CHAPTER GUIDE 1
1.2 AN INTRODUCTION TO NETWORKS IN EDUCATION: RESEARCH AND POLICY 1
1.3 DEFINING NETWORKS IN CONTEXT: STRUCTURE AND CLAIMS 3
1.4 STUDY RATIONALE AND AIMS 4
1.5 CONTRIBUTION OF THE THESIS 4
1.6 STRUCTURE OF THE THESIS 5

2 CONTEXT FOR THE STUDY: NEW YORK CITY PUBLIC SCHOOLS 9
2.1 CHAPTER GUIDE 9
2.2 CHAPTER INTRODUCTION 9
2.3 A HISTORY OF THE DEVELOPMENT OF NETWORKS IN NEW YORK CITY 10
2.3.1 Networks in the context of a broader political agenda 10
2.3.2 Phases of changing school support structures 11
2.3.3 Controversy and dissolution 12
2.4 ASSESSMENTS AND ACCOUNTABILITY 14
2.4.1 Teacher accountability 14
2.4.2 School accountability 15
2.4.3 Network accountability 16
2.5 INITIATIVES AND POLICY CHANGES CONCURRENT WITH NETWORK IMPLEMENTATION 17
2.5.1 Common Core State Standards 17
2.5.2 Union contract 18
2.5.3 Proliferation of charter schools in the NYCDOE 'portfolio' of schools 19
2.6 CHAPTER SUMMARY 19

3 LITERATURE REVIEW 20
3.1 EDUCATIONAL EFFECTIVENESS AND SCHOOL IMPROVEMENT 20
3.1.1 History of EER 20
3.1.2 More recent developments in EER 22
3.1.3 School improvement 26
3.2 LEADERSHIP AND LINKS TO EER 29
3.2.1 Dominant leadership theories 29
3.3 SCHOOL NETWORKING AND COLLABORATION 32
3.3.1 The place of theory in researching networks 32
3.3.2 Studies of networks 33
3.4 CONCLUDING REMARKS: TOWARD A THEORETICAL PERSPECTIVE 35

4 METHODS 38
4.1 CHAPTER GUIDE 38
4.2 AIMS AND RESEARCH QUESTIONS 38
4.3 MIXED METHODS RESEARCH DESIGN 39
4.3.1 Why mixed methods? 39
4.3.2 Classifying the research design 40
4.3.3 Summary of the specific design 41
4.3.4 Remarks on integration 43

4.4 THE ROLE OF THEORY ACROSS STRANDS 45
4.5 QUANTITATIVE DATA, SAMPLES AND METHODS 46
4.5.1 Secondary data 46
4.5.2 Sample 51
4.5.3 Missing data 55
4.5.4 Quantitative analysis: Multilevel modeling 59
4.5.5 Informing the selection of schools for qualitative study 63
4.6 QUALITATIVE DATA, SAMPLE AND METHODS 64
4.6.1 Why case study? 65
4.6.2 About the sample and sampling process 66
4.6.3 Qualitative data collection/sources 68
4.6.4 Qualitative analysis 71
4.7 POSITIONING 73
4.7.1 Theoretical perspective and epistemological stance 73
4.7.2 Personal/professional background 74
4.8 MIXED ANALYSIS: REVISITING INTEGRATION AND SYNERGY 76
4.9 COMMENTS ON VALIDITY 76
4.9.1 Quantitative 76
4.9.2 Qualitative 77
4.9.3 Mixed 77
4.10 COMMENTS ON ACCESS 78
4.10.1 Access to data 78
4.10.2 Access to individuals 79
4.11 ETHICAL CONSIDERATIONS 80
4.12 LIMITATIONS OF THE STUDY 83

5 EXPLORING AND EXPLAINING ACHIEVEMENT VARIATION IN THE NETWORK STRUCTURE FOR SCHOOL SUPPORT 85
5.1 CHAPTER GUIDE 85
5.2 CHAPTER INTRODUCTION 85
5.3 APPROACH TO ANALYSIS 86
5.3.1 Model specifications 91
5.4 DESCRIPTION OF SAMPLE DATA 95
5.5 PARTITIONING VARIANCE 99
5.5.1 Network and school effects 101
5.5.2 District and school effects 101
5.5.3 School effects measured by the ICC 105
5.6 ACCOUNTING FOR STUDENT-LEVEL PREDICTORS: VALUE ADDED MODELS 105
5.6.1 Prior achievement 105
5.6.2 Student background characteristics 110
5.7 ACCOUNTING FOR SCHOOL CONTEXT AND COMPOSITION: CONTEXTUAL VALUE ADDED MODELS 114
5.8 CLASSIFYING SCHOOL PERFORMANCE 118
5.9 CONCLUSIONS 121

6 NETWORK PARTICIPATION: EXPERIENCES AND PERSPECTIVES 124
6.1 CHAPTER GUIDE 124
6.2 CHAPTER INTRODUCTION 124
6.3 NOTES ON THE ANALYTICAL PROCESS 125
6.4 DESCRIPTIONS OF NETWORK ROLES AND SUPPORT OFFERINGS 127
6.4.1 Roles of networks 127
6.4.2 Types of support 130
6.5 NETWORKING WITHIN NETWORKS: CONNECTING WHOM, AND FOR WHAT PURPOSE? 137

6.5.1 Principal to principal *138*
6.5.2 Principal to network staff *141*
6.5.3 Teacher to network staff *143*
6.5.4 Teacher to teacher *143*
6.6 NAVIGATING AND NEGOTIATING NETWORK SUPPORT 144
6.6.1 Choosing and changing: Rationales for CFN affiliation choices *144*
6.6.2 Negotiating support within networks *148*
6.6.3 Perceived value and impact of CFNs and the CFN structure *149*
6.7 PUTTING PERSPECTIVES IN CONTEXT: VIEWS ON EXTERNAL PRESSURES, INITIATIVES, AND POLICIES 153
6.8 CHAPTER SUMMARY 159

7 ACHIEVEMENT IN SCHOOLS AND NETWORKS: INVESTIGATING CHANGE OVER TIME 161
7.1 CHAPTER GUIDE 161
7.2 CHAPTER INTRODUCTION 161
7.3 NOTES ON THE ANALYTICAL APPROACH 162
7.4 MULTILEVEL ANALYSIS RESULTS: 2009-10 TO 2012-13 165
7.4.1 Partitioning variance *165*
7.4.2 Contextualised value added (CVA) models *169*
7.5 STABILITY OVER TIME 177
7.6 PATTERNS OF HIGH, TYPICAL AND LOW SCHOOL PERFORMANCE OVER TIME 180
7.6.1 Patterns over 1-year time spans *181*
7.6.2 Patterns over two-year time spans *185*
7.6.3 Patterns over a three-year time span *187*
7.7 CHAPTER SUMMARY 191

8 DISCUSSION AND CONCLUSIONS 193
8.1 CHAPTER GUIDE 193
8.2 CHAPTER INTRODUCTION 193
8.3 INTEGRATED CONCLUSIONS FROM QUANTITATIVE AND QUALITATIVE EVIDENCE 194
8.4 CONTRIBUTION TO THE RELEVANT LITERATURE 198
8.4.1 Contribution to the literature on CFNs *199*
8.4.2 Contribution to the literature on networks and collaboration in education *202*
8.4.3 Contribution to the literature on educational effectiveness *204*
8.5 STRENGTHS AND LIMITATIONS 205
8.6 DIRECTIONS FOR FUTURE RESEARCH 208
8.7 IMPLICATIONS FOR POLICY 210
REFERENCES 214
APPENDICES 235

LIST OF TABLES

Table 1 Data sources and availability....49
Table 2 Operational definitions of variables....50
Table 3 Final sample size (students, schools, networks) by year....52
Table 4 Descriptive statistics by year....54
Table 5 Numbers of participants by stakeholder group....67
Table 6 Cross-tabulation of schools in the sample by performance category and subject....67
Table 7 Descriptive statistics by category (Literacy, 2011-12)....98
Table 8 Descriptive statistics by category (Mathematics, 2011-12)....99
Table 9 Empty model results, Literacy 2011-12....103
Table 10 Empty model results, Mathematics 2011-12....104
Table 11 Comparison across models, Literacy 2011-12 (ICCs, total variance, and variance explained)....106
Table 12 Comparison across models, Literacy 2011-12 (ICCs, total variance, and variance explained)....106
Table 13 Value added (prior achievement) model results: Literacy, 2011-12....108
Table 14 Value added (prior achievement) model results: Mathematics, 2011-12....109
Table 15 Value added (prior achievement and student background) model results....112
Table 16 Contextualised value added model results: Literacy, 2011-12....115
Table 17 Contextualised value added model results: Mathematics, 2011-12....116
Table 18 Correlations between school effects from empty, value added, and contextual value added models (Stages 1-4)....118
Table 19 Number of schools by performance category in each subject, 2011-12....121
Table 20 Cross-tabulation of schools by performance category and subject 2011-12....121
Table 21 DIC and differences across model structures, 2009-10 to 2012-13....168
Table 22 ICCs for literacy and mathematics models, 2009-10 to 2012-13....169
Table 23 Summary of results for CVA models, 2009-10....172
Table 24 Summary of results for CVA models, 2010-11....173
Table 25 Summary of results for CVA models, 2012-13....174
Table 26 Summary of results for cross-classified CVA literacy model, 2012-13....176
Table 27 Stability of correlations of residuals at the school level: 2009-10 to 2012-13 Literacy, empty models....177
Table 28 Stability of correlations of residuals at the school level: 2009-10 to 2012-13 Mathematics, empty models....178

Table 29 Stability of correlations of residuals at the school level : 2009-10 to 2012-13 Literacy, CVA models 179
Table 30 Stability of correlations of residuals at the school level: 2009-10 to 2012-13 Mathematics, CVA models 179
Table 31 Consistency of correlations between literacy and mathematics residual estimates for schools, empty models 2009-10 to 2012-13 180
Table 32 Consistency of correlations between literacy and mathematics residual estimates for schools, CVA models, 2009-10 to 2012-13 180
Table 33 Counts and percentages of schools categorised as high-, typical- and low performing for each year: 2009-10 to 2012-13 181
Table 34 Cross-tabulation of school performance categories, 2009-10 to 2010-11 (Literacy) 182
Table 35 Cross-tabulation of school performance categories, 2010-11 to 2011-12 (Literacy) 182
Table 36 Cross-tabulation of school performance categories, 2011-12 to 2012-13 (Literacy) 183
Table 37 Cross-tabulation of school performance categories, 2009-10 to 2010-11 (Mathematics) 184
Table 38 Cross-tabulation of school performance categories, 2010-11 to 2011-12 (Mathematics) 184
Table 39 Cross-tabulation of school performance categories, 2011-12 to 2012-13 (Mathematics) 184
Table 40 Cross-tabulation of school performance categories, 2009-10 to 2011-12 (Literacy) 185
Table 41 Cross-tabulation of school performance categories, 2010-11 to 2012-13 (Literacy) 186
Table 42 Cross-tabulation of school performance categories, 2009-10 to 2011-12 (Mathematics) 186
Table 43 Cross-tabulation of school performance categories, 2010-11 to 2012-13 (Mathematics) 186
Table 44 Cross-tabulation of school performance categories, 2009-10 to 2012-13 (Literacy) 187
Table 45 Cross-tabulation of school performance categories, 2009-10 to 2012-13 (Mathematics) 187

Table 46 Cross-tabulation of school performance patterns and network affiliation change, 2009-10 to 2011-12 (Literacy) 188
Table 47 Cross-tabulation of school performance patterns and network affiliation change, 2009-10 to 2011-12 (Mathematics) 189
Table 48 Cross-tabulation of school performance patterns and network affiliation change, 2010-11 to 2012-13 (Literacy) 189
Table 49 Cross-tabulation of school performance patterns and network affiliation change, 2010-11 to 2012-13 (Mathematics) 190
Table 50 Cross-tabulation of school performance patterns and network affiliation change, 2009-10 to 2012-13 (Literacy) 190
Table 51 Cross-tabulation of school performance patterns and network affiliation change, 2009-10 to 2012-13 (Mathematics) 191
Table 52 Frequency and proportion of missing values by variable and year 237
Table 53 Table of missing data patterns by variable (2009-10) 242
Table 54 Table of missing data patterns by variable (2010-11) 242
Table 55 Table of missing data patterns by variable (2011-12) 242
Table 56 Table of missing data patterns by variable (2012-13) 243
Table 57 Little's MCAR test results by year and variable level 244
Table 58 Summary of empty model results, Literacy 2009-10 248
Table 59 Summary of empty model results, Mathematics 2009-10 249
Table 60 Summary of empty model results, Literacy 2010-11 250
Table 61 Summary of empty model results, Mathematics 2010-11 251
Table 62 Summary of empty model results, Literacy 2012-13 252
Table 63 Summary of empty model results, Mathematics 2012-13 253

LIST OF EQUATIONS

Equation 5-1 2-level empty model: Students within schools 91
Equation 5-2 3-level empty model: Students within schools within networks 91
Equation 5-3 3-level empty model: Students within schools within districts 92
Equation 5-4 4-classification empty model: Students within schools, cross-classified within networks and districts 92
Equation 5-5 2-level prior attainment value added model (literacy) 93
Equation 5-6 2-level prior attainment value added model (mathematics) 93
Equation 5-7 2-level prior attainment and student background value added model (mathematics 94
Equation 5-8 2-level prior attainment and student background value added model (mathematics) 94
Equation 5-9 2-level contextual value added model (literacy) 95
Equation 5-10 2-level contextual value added model (mathematics) 95

LIST OF FIGURES

Figure 1 Major components of the ESI framework 28
Figure 2 Procedural diagram 43
Figure 3 Classification diagrams: 2-level, 3-level, and cross-classified models 61
Figure 4 Illustration of high, typically and low performance categories based on residuals and confidence intervals 64
Figure 5 Residuals/confidence intervals, effects of sample schools: Literacy 2011-12 68
Figure 6 Residuals/confidence intervals, effects of sample schools: Mathematics 2011-12 68
Figure 7 Stages of analysis: Gross, value added and contextualised value added models 86
Figure 8 School (shrunken) residuals and confidence intervals by rank: Literacy, 2011-12 119
Figure 9 School (shrunken) residuals and 95% confidence intervals by rank: Mathematics, 2011-12 120
Figure 10 Extent of missing data: Variables, cases and values (2009-10) 235
Figure 11 Extent of missing data: Variables, cases and values (2010-11) 236
Figure 12 Extent of missing data: Variables, cases and values (2011-12) 236
Figure 13 Extent of missing data: Variables, cases and values (2012-13) 236
Figure 14 Missing data patterns (2009-10) 238
Figure 15 Percentage of cases by pattern (2009-10) 239
Figure 16 Missing data patterns (2010-11) 239
Figure 17 Percentage of cases by pattern (2010-11) 240
Figure 18 Missing data patterns (2011-12) 240
Figure 19 Percentage of cases by pattern (2011-12) 240
Figure 21 Percentage of cases by pattern (2012-13) 241
Figure 20 Missing data patterns (2012-13) 241
Figure 22 Normal probability plot (L2), 2009-10 Literacy 254
Figure 23 Normal probability plot (L2), 2009-10 Mathematics 254
Figure 24 Normal probability plot (L2), 2010-11 Literacy 254
Figure 25 Normal probability plot (L2), 2010-11 Mathematics 254
Figure 26 Normal probability plot (L2), 2011-12 Literacy 255
Figure 27 Normal probability plot (L2), 2011-12 Mathematics 255
Figure 28 Normal probability plot (L2), 2012-13 Literacy 255
Figure 29 Normal probability plot (L2), 2012-13 Mathematics 255

LIST OF APPENDICES

Appendix 1 Missing data analysis 235
Appendix 2 Principal consent form -- access to school 245
Appendix 3 Teacher information letter 246
Appendix 4 Interview consent form 247
Appendix 5 Summaries of empty model results (2009-10, 2010-11, 2012-13) 248
Appendix 6 Normal probability plots for model residuals: Final CVA models 254

LIST OF ABBREVIATIONS

CCSS(I)	Common Core State Standards (Initiative)
CFN	Children First Network(s)
CVA	Contextual value added
DASI	Dynamic Approach to School Improvement
DBN	District-Borough-Number
DIC	Deviance Information Criterion
DOE	Department of Education
EER	Educational Effectiveness Research
ELA	English Language Arts
ESI(P)	Effective School Improvement (Project)
ESO	Empowerment Support Organisation
FSM	Free school meals
ICC	Intraclass correlation
IRB	Internal Review Board
ISIP	International School Improvement Project
LSO	Learning Support Organisation
MAR	Missing at random
MCAR	Missing completely at random
MCFA	Multilevel Confirmatory Factor Analysis
MCMC	Markov Chain Monte Carlo
MEFA	Multilevel Exploratory Factor Analysis
MM	Mixed methods
NLC	Networked Learning Communities
NMAR	Not missing at random
NYCDOE	New York City Department of Education
PMM	Portfolio Management Model
PSO	Partnership Support Organisation
RPSG	Research and Policy Support Group
SEM	Structural Equation Modeling
SI	School Improvement
SSO	School Support Organisation
UK	United Kingdom
US	United States

1 INTRODUCTION

1.1 CHAPTER GUIDE

This chapter is intended to:

- Introduce the field and background for the study
- Describe the structure and definition of networks in context
- Present the rationale and purpose of the study
- Explain the proposed structure of the thesis, addressing how chapters contribute and connect to the main argument

1.2 AN INTRODUCTION TO NETWORKS IN EDUCATION: RESEARCH AND POLICY

Networking and collaboration have come into vogue in education on a global scale. Numerous policies have introduced school networks and collaborative initiatives within and between institutions as a means to promote school improvement, relying on the assumption that collaboration is inherently beneficial as well as suggesting that bottom-up approaches to stimulate change are better than top-down policies. Illustrative examples of inter-school networks in the UK can be seen in Federations (Lindsay et al., 2007), academy chains (Salokangas & Chapman, 2014), and Networked Learning Communities (Katz & Earl, 2010; Sammons et al., 2007), while in the US such networks have often arisen out of broader initiatives like the Annenberg Challenge (see Smylie & Wenzel, 2003; Wohlstetter, Malloy, Chau, & Polhemus, 2003) and Comprehensive School Reform (Glazer & Peurach, 2013). A common underlying principle seems to be that with schools subject to demanding standards and challenging circumstances, they are unlikely to achieve and sustain improvement alone; networks, then, are hoped to have the potential to help schools navigate challenges, manage change, and sustain improvement collectively and collaboratively.

Because of the policy emphasis on and prevalence of networks in education, they have also become the focus of a growing body of literature in educational research. The coherence of the empirical and theoretical knowledge base, however, is challenged by the fact that definitions of networks vary between authors, just as structures and purposes of

networking and collaboration differ between settings. A "network" may refer to connections within or between schools and other organisations, and may be defined in terms of either or both form or purpose (e.g. a minimum number of organisations, or a cohesive set of common goals); similarly, "collaboration" may encompass varying levels of formality, involvement and commitment (Shinners, 2001; Hanford et al., 1997). In short, as noted by Leonard and Leonard (2001), "Defining collaboration may be as difficult as achieving it" (p.387). This definitional challenge is revisited in later chapter, but for the sake of clarity, this research adopts the definition that networks are "at least two organisations working together for a common purpose for at least some of the time" (Muijs, West, & Ainscow, 2010, p.6). This definition provides considerable flexibility, but simultaneously bounds the scope of the relevant literature by drawing a distinction between the types of inter-organisational networks discussed in this thesis, and less closely related forms of intra-school or interpersonal networks.

Proceeding with this definition, the majority of research on networks in education addresses some formulation of one or more of the following questions:

- What happens in networks?
- What outcomes can be attributed to networks?
- What makes some networks more or less effective at achieving desired outcomes?

The existing evidence base associated with each of the above questions varies in quality and conclusions. Numerous studies provide rich descriptions and participants' accounts of the processes and interactions within networks, but evidence of networks' impact and effectiveness is more limited and often conflicting, particularly with regard to student achievement. Some studies have found significant effects, though only in particular circumstances or for particular types of networks (e.g. Chapman & Muijs, 2013; Katz & Earl, 2010), while others have found little to no effect (e.g. Smylie & Wenzel, 2003; Sammons & Mujtaba, 2007). In settings such as the US and UK, where student achievement is prioritized for accountability purposes, this constitutes a striking disparity between the research evidence base and policy decisions relevant to networking and collaboration.

1.3 DEFINING NETWORKS IN CONTEXT: STRUCTURE AND CLAIMS

This thesis focuses on networks of schools in New York City, known as Children First Networks (CFN). These networks constitute a highly formalised structure; each public school has been required to affiliate with one of approximately 60 networks since 2010, and each network is comprised of a distinct organisational entity with its own staff (network leaders and various specialists and coaches) linked with up to 35 schools. Although it is mandatory for schools to have a network affiliation, they may choose or change this affiliation annually. Networks are further grouped within and overseen by "clusters". Most clusters are internal to the New York City Department of Education (NYCDOE), while one is made up of external non-profit and university organisations called Partnership Support Organizations (PSOs), contracted by the NYCDOE to provide school support.

A set of explicit claims frames the intended purpose and impact of these networks. According to the NYCDOE, the network structure for school support is "dynamic and responsive to individual school and community needs" (NYCDOE, 2014), ultimately allowing schools to "accelerate student achievement" (NYCDOE, 2012). The support offered by networks is designed to be comprehensive, including providing professional development and training for teachers and school leaders, distributing resources for curriculum and instruction, assisting schools with operations (e.g. pertaining to budgeting, hiring, and facilities), and facilitating collaboration between schools.

With respect to accountability in the NYCDOE, the location of networks within the existing hierarchy is somewhat complex. Although networks provide support for schools, schools are not accountable to the network entity; rather, they are accountable to superintendents based in geographical districts distinct from networks. On the other hand, networks are accountable to the NYCDOE for the performance of their affiliated schools, and are evaluated annually based on this as well as principals' satisfaction with their networks.

While the claims about these networks and their general function are clear, the specific processes at work within them are not transparent. Little rigorous independent research has been conducted to examine these processes, and even less to investigate the effects of networks in context. Empirical research to date includes exploratory studies by Wohlstetter, Smith, and Gallagher (2013) and Wohlstetter, Houston and Buck (2015),

which used interviews with network staff (and several principals, in the former study) to examine organisational learning in networks and the role of networks in the implementation of a specific policy. An additional study commissioned by the NYCDOE aimed to identify the "core values" of the network structure for school support and provide policy recommendations to improve upon it (Parthenon Group, 2013), but the methods and data used to arrive at these findings are reported somewhat vaguely, and a further study by the Annenberg Institute (Fruchter et al., 2015) investigated network effects on achievement but failed to fully account for the underlying multilevel structure (the discussion in Chapter 8 further elaborates on this point). The present study is the first of its kind focusing on networks in the New York City context to combine rigorous quantitative approaches with qualitative inquiry to account for effects on achievement as well as stakeholder perspectives and understandings.

1.4 STUDY RATIONALE AND AIMS

This research takes as its point of departure the explicit claims and implicit assumptions presented above about New York City's networks, their measurable effects on student achievement, and the processes and interactions through which they support schools and school improvement. Thus, the aims of this study are to investigate whether there is in fact an effect of networks on student achievement, to explore the processes and interactions within networks from multiple perspectives, and to draw connections between these two aspects, ultimately enhancing understanding of the role of networks with regard to school effectiveness and improvement.

1.5 CONTRIBUTION OF THE THESIS

Through the detailed study of networks in a specific and under-researched context, this research makes an original contribution to the existing literature on networking and collaboration in education by using a rigorous mixed methods approach to connect processes, perspectives, and outcomes. There is some precedent for the use of multilevel modelling to analyse the effects of networking on student outcomes (Chapman & Muijs, 2014), and case study approaches are common in school networks research (Atkinson et al., 2007), but integration of the two in a single study as is done in the research presented in this thesis is relatively rare with few exceptions (e.g. Chapman & Muijs, 2013). In addition, this research contributes to the educational effectiveness knowledge base by

applying well-established methods for studying school effectiveness to a different set of organisational structures including students nested within schools, and schools nested (cross-classified) within both networks and geographical districts.

From a practical perspective, findings presented in this thesis have the potential to inform policy relevant to networking and collaboration in New York City. Results pertaining to effects on student achievement may temper claims and expectations about the benefits of networks in their current form, while qualitative findings offer insight into the realities of implementation, as well as the experiences and perspectives of practitioners, at the school level. Generalisations to other contexts must be made cautiously; however, given that networks in New York City reflect a broader trend in education, this research may provide useful implications for policymakers considering network initiatives in other contexts, while also illuminating participants' experiences and providing knowledge relevant to practitioners.

1.6 STRUCTURE OF THE THESIS

This section describes each proposed chapter to be included in the final thesis, and explains the connections between chapters as well as their contributions to the thesis as a whole.

The first four chapters present the background, context, theoretical foundations and methodological approach of the study. The next three chapters present findings and interpretations from quantitative and qualitative strands of inquiry, alternating between the two and demonstrating how each builds upon and feeds into the other. The final chapter draws together and synthesises the main findings, integrating understandings from qualitative and quantitative strands, and analyses the implications of these against the backdrop of the context and relevant literature.

Chapter 1 presents a brief introduction to the background, context, and main argument of the thesis. Sections include a concise account of policy and research on networking and collaboration in education, a detailed description of networks in context, a statement of the rationale and aims of the study, and an explanation of the original contribution of the thesis to the relevant research literature as well as to policy. The final section of the introduction presents the structure of the thesis, summarising the contents of each chapter and explaining how they link together.

Chapter 2 provides a descriptive account of relevant aspects of the New York City context for public education, building on the description of networks given in Chapter 1. Sections cover a combination of policy, research, and narrative of key changes in the system from 2002-present. Sections are also included to introduce the reader to relevant details about assessment and middle schools in context, as a basis for better understanding the study's design and findings. The purpose of the chapter is to provide the reader with an understanding of public education in context, of the policies and politics that led to the formation of networks, and of the relevant constellation of initiatives implemented over the same time period during which networks are studied in this research.

Chapter 3 presents a review of relevant literature in the areas of educational effectiveness and networking and collaboration, as this study lies at an intersection between these. Particular attention is given to methodological developments in educational effectiveness research (EER) that inform the methodology for the present study, and the relationship between EER and school effectiveness is analysed to form a bridge between fields from EER to networks as a school improvement device. Following that, a critical analysis of the literature on networks and collaboration in education is presented, focusing on the methodological approaches used to study networks, empirical results concerning effects and impact of networks, and theoretical perspectives on networking and collaboration that, together with an educational effectiveness perspective, form the theoretical frameworks for this study. Gaps in the literature are highlighted, and the contribution of the thesis is revisited in light of the existing literature.

Chapter 4 offers a detailed account of the research design and methodology of the study, linking back to the theory discussed in Chapter 3 where appropriate. Beginning with the aims and research questions, the chapter proceeds with a description of the overall mixed methods design, followed by details of the quantitative approach including data sources, sample description, data collection, and analysis, with corresponding sections for the qualitative strand. An explicit treatment of the role of and approach to integration is given, and issues of validity and reliability are considered from qualitative, quantitative, and mixed perspectives. In addition to a clear presentation and justification of the methods used, the contents of the chapter also address the positioning of the researcher, and the ways in which the research design has developed and changed over the course of the project. Final sections discuss access, ethics, and limitations of the study. In the context of the thesis as a whole, this chapter sets the stage for the reader to interpret the findings

presented in the chapters that follow, tracing a clear progression from methodology and methods to results.

Chapter 5 presents the findings from initial quantitative analyses. Connecting back to the aims and research questions given in Chapter 4, the first sections develop a more detailed description of the analytical process, including model specifications and issues considered at each step of model fitting (model fit, significance of coefficients, intra-class correlation, variance explained, and consistency of results across subjects). Findings are presented and interpreted for four sequential stages of model fitting: empty (unadjusted) models, value added models adjusting for student prior achievement, value added models correcting for additional student predictors, and further-developed CVA models controlling for school composition. Besides providing answers to a subset of the questions given in Chapter 4, this chapter demonstrates how results were used to define categories of schools (high-, typically-, and low-performing) to inform sampling for the qualitative strand of the study. The results presented in this chapter also link back to the discussion of empirical literature in Chapter 3, contributing to the broader evidence base concerning effects of networking and collaboration on student achievement.

Chapter 6 presents results from interviews with network staff, principals, and teachers. The analytical process discussed in Chapter 4 is further developed here, with particular attention to the co-construction of meaning with participants. This is followed by sections defined according to key themes emerging from the analysis of interviews. These themes address definitions and understandings of the roles and impact of, as well as the nature of participation in, networks, the existence and nature of informal networks shaping as well as emerging from the network structure, the processes and priorities driving network affiliation, and the external pressures and initiatives affecting participants' perspectives. This chapter provides insight into the processes and interactions within schools and networks that are otherwise invisible, addressing a key subset of the questions presented in Chapter 4, and providing possible explanations for some of the results presented in previous findings chapters.

Chapter 7 builds on the analysis and results presented in Chapter 5, extending the previous models to consider achievement over time. First, the CVA models used to inform sampling are applied to additional years and results are compared over time with regard to network and school effects. Second, stability is investigated via correlations between

school value added effects between different years. Third, school performance categories based on CVA effects as defined in Chapter 5 are used to investigate patterns of network affiliation and change, and possible relationships of these to school performance. Results are linked to findings from interviews presented in Chapter 6 to suggest explanations for the statistical results as well as to findings from Chapter 5 to identify how the new findings presented in Chapter 7 build on, are explained by, or are explained by those presented in earlier sections. The results in this chapter contribute to the thesis as a whole by providing a more complete and detailed analysis of effectiveness and change in schools and in the network structure, accounting for a longer time period and addressing research questions not yet dealt with in earlier chapters.

The final chapter of the thesis, **Chapter 8**, synthesises and integrates the main findings from previous chapters. Different understandings of networks in context, their roles and theis effectiveness are interpreted through the lens of the quantitative findings, from the multiple perspectives of participants, and through a combination of both of these. The ways in which these converge, diverge or complement each other are discussed. Conclusions are linked back to the relevant literature on networking and collaboration and critically analysed alongside previous research and helping to inform gaps in the knowledge base and giving particular attention to conflicting findings and the possible reasons for them. A variety of directions for future research are identified and explained, and based on key findings, conclusions are drawn with respect to policy and practice in context.

2 CONTEXT FOR THE STUDY: NEW YORK CITY PUBLIC SCHOOLS

2.1 CHAPTER GUIDE

This chapter is intended to:

- Provide the reader with an understanding of the context for the present study
- Provide a basis for understanding some of the key terms particular to the New York City policy and practice context

2.2 CHAPTER INTRODUCTION

New York City is home to the largest public school system in the US, with (as of 2013) approximately 1,700 schools, 75,000 teachers, and 1.1 million students. The city is made up of five boroughs: Manhattan, Brooklyn, Queens, the Bronx, and Staten Island. Although it is unique in terms of sheer size, New York City's schools face challenges common to many urban US school districts. These include sizeable equity gaps in attainment between ethnic and socioeconomic groups, funding shortages, and tensions between various political agendas (Hunter & Brown, Frank, Brown, 2002).

In order to address these challenges, New York City has undergone a rapid progression of numerous and overlapping reforms over roughly the last two decades. This chapter contextualises the study of Children First Networks (CFNs) within this period of rapid and complex educational change. The first section below addresses how networks developed, and the policies and politics that drove their formation. The next section introduces and describes other key reforms separate temporally overlapping with CFN implementation, prioritising those initiatives and policy changes that emerged as important to contextualise participants' perspectives in the qualitative findings of this study. After that, some background on assessments and accountability is presented to facilitate understanding of the outcome measures used in this study as well as their role in accountability systems in the NYCDOE and their influence on judgments of school and network performance in context.

2.3 A HISTORY OF THE DEVELOPMENT OF NETWORKS IN NEW YORK CITY

2002 marked a turning point in education policy in New York City. Until this time, governance centred on a publicly-elected school board, and schools were grouped into geographically defined districts. Students attended schools according to their place of residence, with the exception of selective middle and high schools with admission determined by scores on standardized examinations. The election of a new mayor in 2002 began a period of sweeping change that restructured the New York City school system and marked an era of dramatic policy reform.

2.3.1 Networks in the context of a broader political agenda

When Michael Bloomberg became Mayor in 2002, one of his first actions was to place the entire school system under mayoral control and reorganize what was then the New York City Board of Education (now the New York City Department of Education, or NYCDOE). The elected school board was replaced by a Chancellor decided by mayoral appointment (initially Joel Klein, a key actor in the design and implementation of policies and reforms described below), and greater authority was assigned to the central office of the NYCDOE (Kelleher, 2014a). This newly centralized authority facilitated the launch of the Children First reform agenda in 2003, which brought about comprehensive restructuring over several years based on driving principles of leadership, empowerment and accountability in schools (NYCDOE, 2008).

There are three major components of the Children First reforms. First, beginning in 2002, low-performing comprehensive neighborhood high schools have been dissolved and replaced by many newly-founded small schools (Iatarola, Schwartz, Stiefel, & Chellman, 2008), often with particular themes (such as the Academy for Careers in Television and Film, or the Bronx Aerospace High School). With this was initiated a market-based system of school choice, with a lottery matching students to schools based partly on student and school preference rankings; this was intended to promote equity and choice, but in practice the choices of low-performing and special needs students are limited (Jessen, 2013; Nathanson, Corcoran, & Baker-Smith, 2013). Second, there was a gradual transition to school-based management, shifting decision-making and budget allocation from the district to individual principals. This was piloted in a selection of schools in 2004 and then expanded to all schools in 2007 (Gyurko & Henig, 2010). Third, School Progress Reports were introduced in 2006, assigning schools an annual letter grade based on school

environment, student performance, and student progress (for high schools, this is measured in terms of credit accumulation and state examination results). New accountability systems brought increased transparency, as well as consequences such as the threat of closure for consistently under-performing schools (Corcoran & Pai, 2013).

Together, these reforms set the foundation for implementation of the Portfolio Management Model (PMM), a strategy based on the work of (P. T. Hill, Pierce, & Guthrie, 1997). PMM varies between settings but has three defining features (Bulkley, 2010):

1) the creation of autonomous new schools and an emphasis on school choice based on students' interests and needs;
2) clear and rigorous accountability systems; and
3) district responsibility for closing schools and ending partnerships that fail to meet accountability standards.

By 2007, New York City had adopted the PMM system-wide. Previously, schools were organised according to geographical districts and then regions, and received support (e.g. professional development) and mandates (e.g. with respect to staffing and budgeting) from geographically-based authorities (Hemphill & Nauer, 2010). Within this model, as part of the effort to increase the autonomy and decision-making authority of principals, networks emerged as a new structure for supporting schools. At the same time, schools were directly accountable to their district superintendents, who held the authority to hire and fire principals. By contrast, networks were not located above schools in the accountability hierarchy; they were intended to provide many of the same supports for schools previously associated with districts, but schools remained accountable to their district superintendents. Principals were given more autonomy and responsibility for budget decisions and staffing, and the freedom to choose and change networks (making networks essentially accountable to schools rather than the other way around) (The Parthenon Group, 2013).

2.3.2 Phases of changing school support structures

Children First Networks did not take shape overnight. The emergence of networks may be traced back to a pilot initiative called the Autonomy Zone, which began in 2004 with 29 participating schools (3 of which were charter schools) and placed decision-making power at the school level with regard to hiring and professional development selection (Kelleher, 2014a). This was expanded to over 300 schools and rebranded as the Empowerment Zone

over the next two years, at which point the beginnings of non-geographically-defined networks were apparent, as the Empowerment Zone was organised into 14 networks of over 20 schools each, with these groupings defined by principals' preferences. These networks each featured an organisational structure involving a network leader and small core team with a variety of instructional, management, assessment, and other expertise (Nadelstern, 2012).

The scale up of networks occurred over the several years that followed. The regional accountability structure gave way to a new model for school support in which schools could choose to affiliate with Empowerment Support Organisations (ESOs) involving an agreement offering principals greater autonomy while requiring them to sign performance contracts with achievement targets, Partnership Support Organisations (PSOs) managed by external nonprofit and university organisations, and Learning Support Organisations (LSOs) which served a similar function to the previous regional offices overseeing schools but were not geographically bounded (Kelleher, 2014a; Nadelstern, 2012). This structure of School Support Organisations (SSOs) had some features that carried over to what later became the CFN network structure, specifically principals' ability to choose between different support organisations and the emergence of a delineation between organisations responsible for providing support and those responsible for accountability (as the role of SSOs with regard to school accountability was not consistent or clearly defined).

In 2009, the NYCDOE began transitioning from SSOs to networks. By 2010-11, the CFN network structure on which this thesis focuses had been fully implemented. PSOs led by external nonprofit and university partners were still a feature of the school support structure and led a selection of networks, and networks were tasked with providing their affiliated schools with a broad range of instructional, operational, and other types of support, while schools remained directly accountable to geographically-based district authorities (The Parthenon Group, 2013). All schools were required to have a network affiliation, but could choose or change this affiliation from year to year. The network structure overall has been described previously in the Introduction (Chapter 1) and so is not repeated here. This structure remained fairly consistent from 2010-11 to 2015-16.

2.3.3 Controversy and dissolution

CFNs represented the autonomy and choice central to the Portfolio Management Model. Proponents of networks argued that they were more cost-effective than previous

organisational structures for support, they offered tailored support for individual schools, and they offered opportunities for knowledge sharing and collaboration across a variety of schools not limited by district boundaries (Kelleher, 2014b; The Parthenon Group, 2013).

On the other hand, opponents cited a lack of transparency and accountability as fundamental flaws of the CFN model for school support, and drew attention to the fact that not being geographically bounded meant that it was a challenge for network staff to travel to schools and vice versa. Just as the implementation of networks was posited above to be values-based, opposition to CFNs often reflected a set of political and philosophical convictions, so that the debate at times took on a combative or accusatory tenor, reflected in a public speech by the president of the teachers' union:

> "Fundamentally, there's a disconnect. From both an educational and operational standpoint, the DOE concerns itself with its own needs and not the needs of the schools. The DOE does not directly interface with the schools; issues are bounced around between the schools and the networks and Tweed. Worse, there's no direct chain of command between the DOE and the networks. We understand some key department heads have no authority over their network counterparts. Decision-making is cumbersome, confusing and takes too long. When there's the inevitable breakdown, Tweed deflects blame; it's the principal's fault, or it's the network's fault. This is not accidental; they choose to operate this way so they won't be held responsible. Either we have a system of networks that are fully empowered or we have the DOE; we don't need both." (Mulgrew, 2012)

A change in mayoral administration in 2013 marked the end of the "Bloomberg-Klein" era (as it is commonly referred to by New York City educators). The new mayor, Bill De Blasio appointed a new Chancellor, Carmen Fariña, whose professional background (which working as a teacher, principal, superintendent, and as Deputy Chancellor for Teaching and Learning) (NYCDOE, 2016a) was considerably different from her predecessors, all of whom either had backgrounds mainly in sectors other than education (including business and law) and two of whom required waivers to take up the Chancellor position due to a lack of educational administration experience (Hernandez & Otterman, 2010; Phillips, 2011). The new administration was openly in opposition to the reforms of the previous administration, and expressed an intention to reverse many of the policies instituted under the leadership of Mayor Bloomberg and Chancellor Klein (Layton & Chandler, 2013).

Ultimately, CFNs proved to be one of the initiatives that the De Blasio-Fariña administration set out to undo. In 2015-16, plans to dissolve these networks by the end of that academic year were announced, replacing CFNs with a system giving more authority to district superintendents and establishing seven geographically based Borough Field Support Centers offering similar forms of school support to those provided by CFNs. In this new structure, principal autonomy and independence was to be preserved "except in cases of struggling schools". PSOs were kept in place as a source of resources and guidance for schools, but were made more directly accountable to superintendents (NYCDOE, 2015). The dissolution of CFNs was controversial in its own right, as advocates of the network structure questioned the policy change, with some suggesting that this was being done "without any significant rationale" (Shapiro, 2015).

These developments occurred after data collection and analysis for the present study had been completed. However, knowledge of these later policy developments is provided here to facilitate the reader's understanding of the broader policy context in the setting in which this research took place.

2.4 ASSESSMENTS AND ACCOUNTABILITY

Accountability was a focus of the education reforms under the Bloomberg administration, and a core value of the PMM (Bulkley, 2010; P. T. Hill et al., 1997). Student scores on New York State Assessments in English Language Arts (literacy) and Mathematics, which are also used as outcomes in the present study, were at the heart of NYCDOE accountability protocols, and constituted a heavily-weighted factor in evaluations of performance at all levels of the school system. The sections below outline how networks, schools and teachers were evaluated annually in New York City during the time period covered in this research.

2.4.1 Teacher accountability

Policy governing teacher evaluations in New York City has undergone numerous changes during the period from 2009-10 to present. State law instituted in 2010 mandated annual professional performance reviews for teachers, with composite scores and ratings of "highly effective", "effective", "developing", or "ineffective" calculated from the following components (The University of the State of New York, 2014):

- 40% measures of student achievement and growth based on a combination of student state assessment scores and locally-selected measures of achievement, and
- 60% other measures of effectiveness offering evidence of proficiency with regard to state teaching standards.

The latter 60% was required to include multiple measures, including a minimum of two annual observations of classroom practice, and could such additional components as peer observations, independent external evaluator observation, parent or student feedback surveys, or structured review of teaching artifacts or lesson plans. Rubrics used for observations ratings were required to be one of a selection approved by the state (NYSED, 2013). In New York City, the Danielson Rubric was widely used; this rubric, which was intended to synthesise empirical and theoretical research to identify aspects of teacher responsibilities that promote student learning, included 22 components clustered within the 4 domains of "Planning and preparation", "The classroom environment", "Instruction", and "Professional responsibilities" (Danielson, 2007).

The aim of this brief description of some key aspects of teacher accountability in context is not to give a comprehensive overview of the relevant policies, but to acquaint the reader with the language of policy and practice surrounding teacher evaluations so that terminology arising in quotes from interviews is understood later in the thesis.

2.4.2 School accountability

During the time period included in this study, there were three types of reports used to evaluate all New York City public schools. These included (NYCDOE, 2010):

- Progress Reports Grades: Each school was rated "A" through "F" (from highest to lowest performance) annually based on a composite measure of student progress[1], a

[1]"Student progress" on progress reports was defined as the sum of median adjusted growth percentiles for all students and for the lowest-performing third of students in literacy and mathematics, where growth percentile for each student was calculated as the percentage of students with the same score on the previous year's state assessment who scored the same or lower on the current year's test (NYCDOE, 2009).

measure of student performance[2], a measure of school environment[3], and a measure termed "closing the achievement gap"[4].

- Quality Review scores: Quality Reviews were inspections conducted by experienced educators from outside of a school (not necessarily every year, depending on a school's rating), rated according to a rubric measuring various dimensions of schools' planning and decision-making, data-collection and analysis, goal-setting, leadership development and professional collaboration, and evaluation and monitoring of progress towards established goals. Quality Review outcomes consisted of an overall rating on a four-point scale from "Well developed" to "Under-developed".
- New York State Annual Report Cards: Based on a combination of student demographic data, student assessment results, attendance data, class size, enrolment, suspensions, staff counts, teacher qualifications and turnover, state report cards reported each school's accountability status under state law.

2.4.3 Network accountability

Like schools, CFNs were rated annually. Network ratings consisted of a score from 1 to 4 (from lowest to highest) combining aggregated measures of the Quality Review scores and Progress Report grades of affiliated schools, qualitative evaluations by teams from the higher "cluster" level of the network structure, and principal satisfaction surveys. Low ratings for two years or a low number of affiliated schools would result in the closure of a particular network (Kelleher, 2014b).

Network ratings were not made consistently public. A spreadsheet of 2011-12 results was posted on the NYCDOE website, but when ratings for other years of CFN implementation were requested from the NYCDOE Research and Policy Support Group (RPSG) as part of the data request for the present study, the response was that no records of these ratings were available. The document showing performance ratings for 2011-12 did not show overall ratings for each network, but listed Progress Report average percentiles and

[2] "Student performance" consisted of the sum of four measures: number of students scoring above a cutoff score for proficiency in literacy, number of students scoring above a cutoff score for proficiency in mathematics, and average proficiency ratings for all students in literacy and mathematics (Ibid.).

[3] "School environment" consisted of a sum of four metrics including school survey scores based on parent, teacher, and student responses and a directly-measured attendance rate (Ibid.).

[4] This awarded extra credit to schools achieving higher-than-expected outcomes for "high-need" groups of students (Ibid.).

resulting network-level quartiles and Quality Review percentiles and network-level quartiles, and percentages of principals in each network who were "very satisfied", "satisfied", "dissatisfied", or "very dissatisfied" with network support. This document also showed that principal satisfaction surveys consisted of a single question, "How satisfied are you with the overall quality of support provided by your network?" (NYCDOE, 2012).

2.5 INITIATIVES AND POLICY CHANGES CONCURRENT WITH NETWORK IMPLEMENTATION

Educational policies and initiatives do not occur in a vacuum, rather, for any given policy shift or newly implemented strategy, there are likely at least some and possibly many additional initiatives or policy changes happening simultaneously. Children First Networks are no exception. During the time period considered (2009-10 to 2012-13), there were major changes in the New York City school system spanning from dramatic shifts in curriculum and assessment to a transition between political administrations with markedly different values and priorities with regard to education. The sections below are by no means an exhaustive list of every reform and initiative that took place over the time period of interest. Instead, they describe changes in the school system that emerged as important and relevant based on the perspectives of interview participants in this study. Thus, the descriptions of changes given below provide the reader with background knowledge to facilitate a fuller understanding of qualitative findings, particularly those regarding participants' views on external pressures and initiatives.

2.5.1 Common Core State Standards

In 2009, the National Governors Association (NGA), which consists of the governors of US states, territories and commonwealths and acts as an intermediary organisation between the federal and state governments, initiated the development of a set of standards intended to promote college and career readiness, equity, and competitiveness in the global economy for all students (CCSSI, 2015). These standards, referred to as the Common Core State Standards, were released in their final form in June, 2010, and are intended to define a set of learning goals for what students must know and be able to do by the end of each grade in school that is consistent at the national level (CCSSI, 2016). Since the release of the standards, four territories and forty-three states have adopted the standards, including New York (NGA, 2015).

The rollout of these standards involved pervasive changes to curriculum and assessments in New York City. Some notable changes included greater emphasis on nonfiction texts in English classes and substantive changes to learning goals in mathematics. These changes were not without controversy. Opponents to the Common Core suggested that rollout was rushed and teachers had inadequate materials and support to fully implement the standards in time for new assessments aligned to these standards (Freedberg, 2016), while supporters insist that although the Common Core did lead to an initial drop in achievement, this initiative will ultimately improve education (New York Times Editorial Board, 2014). Arguments on both sides lack a trustworthy evidence base at present, leaving the Common Core State Standards as a point of contention in public opinion.

As providers of school support, CFNs played a key role in the rollout of the Common Core State Standards in New York City schools during the time period included in the present study (Wohlstetter et al., 2015), making this initiative a particularly relevant aspect of the policy context for this research.

2.5.2 Union contract

The new mayoral administration was widely viewed as having a positive relationship with the teachers' union in New York City, in contrast to the previous administration's openly hostile relationship with the union (Dawsey, 2016). One of the early developments of the De Blasio administration was a contract negotiation that ended with a 9-year agreement, marking the first time a union contract had been in place since 2009. The contract applied retroactively to the years during which contract negotiations had remained unresolved under the previous administration, and involved items including a retroactive wage increase, changes to teacher evaluation rubrics, dedicated time in schools for professional development and parent engagement, changes to policy relevant to disciplining or removing teachers for misconduct or ineffectiveness (Press Office of the City of New York, 2014). Contract items had strong implications for the policy change at the system level of education in New York City as well as for the day-to-day work of school-based staff, and as the timing of contract negotiations overlapped with the timing of qualitative data collection for the present study, the union contract and the reactions of various stakeholders to it form a significant part of the contextual backdrop for this research.

2.5.3 Proliferation of charter schools in the NYCDOE 'portfolio' of schools

The proliferation of charter schools was a feature of the PMM and its emphasis on school choice (Bulkley, 2010). These schools, which are publicly funded but are not subject to many of the same regulations and accountability structures as public schools, grew in number from seventeen at the beginning of the Bloomberg administration to a hundred by 2012 (Winters, 2012). The principles underpinning the PMM inherently put these schools in direct competition with their public counterparts for student enrolment. Charters are a highly polarising issue in terms of public opinion. Advocates cite higher achievement compared to public schools. Opponents voice concerns including non-unionised teachers and students being "pushed out" for low performance, although there is some research evidence to dispute the latter point (Winters, 2015). Regardless of their possible merits or drawbacks, charter schools form an integral part of the educational landscape in New York City, and while these schools fall outside the scope of the present study, they emerged as a relevant contextual consideration in the qualitative findings presented later in this thesis.

2.6 CHAPTER SUMMARY

The networks that form the focus of this thesis were the product of a dramatic and often controversial set of reforms that took place in New York City beginning in 2002. This chapter provided a narrative account of how and why CFNs evolved and located this initiative within a broader political and policy context. This background helps to contextualise the research findings, and provides the reader with a primer on the language and policy context to facilitate understanding of participants' perspectives presented in Chapter 6.

3 LITERATURE REVIEW

The following sections provide an overview of theory and research in three fields relevant to the current study: educational effectiveness and school improvement, leadership, and school networks. This review aims to accomplish three goals: first, to contextualize this study against the backdrop of prior research; second, to identify gaps in the literature and areas for further development; and third, to introduce a theoretical perspective for this research.

3.1 EDUCATIONAL EFFECTIVENESS AND SCHOOL IMPROVEMENT

3.1.1 History of EER

The origins of Educational Effectiveness Research (EER) trace back to the 1960's and early 70s, when large-scale studies in the US, concerned with the relationship between schooling and equity of opportunity, found that after accounting for student background characteristics, differences between schools accounted for little of the variance in student outcomes (Coleman et al., 1966; Jencks et al., 1972; Hauser, 1971).

These studies were criticized for methodological flaws in their statistical interpretations, as well as an overemphasis on cognitive outcomes and lack of measures of educational attainment (e.g. Averch, 1971; Madaus, Kellagham, Rakow, & King, 1979). There was also an ideologically based reaction from researchers who were convinced that schools must make a difference. This conviction underpinned the emergence of studies in the US aimed at describing exceptionally well-performing urban schools serving disadvantaged students (e.g. Weber, 1971). Out of this trend was born what Teddlie and Reynolds (2000) describe as the 'equity ideal' (p.10), a guiding principle that led researchers to try to identify the characteristics that made schools effective for typically underperforming students.

Studies conducted by Edmonds (1979) in the US and Rutter *et. al.* (1979) in England are widely regarded as the first explicitly designed school effectiveness studies. Edmonds synthesized findings from his own and other studies (e.g. Brophy & Good, 1970; Weber, 1971) into a 'five-factor model' for effective schools with

the following components: strong administrative leadership, high academic expectations, a focus on basic skills acquisition, a safe and orderly school climate, and frequent monitoring of students' progress. Meanwhile, Rutter's (1979) cohort study suggested that differences in delinquency, attendance, behavior problems, and achievement between schools might be associated with 'ethos', made up of a cluster of factors including cohesion, strong academic emphasis, high expectations, positive teacher attitudes towards students, and shared values and norms. These studies were criticized on methodological grounds and for lacking a theoretical foundation, and in particular Edmonds and others in the US concerned with creating 'effective schools for the urban poor' were accused of sampling bias, unfounded claims of causality, and narrowly-defined conceptions of effectiveness (e.g. Cuban, 1984). It was precisely these criticisms of pioneering effective schools research that paved the way for future developments in EER, not least a more methodical approach to context.

Context studies in the 1980s (e.g. Hallinger & Murphy, 1986; Teddlie & Stringfield, 1985) expanded the search for factors related to effectiveness to a broader range of schools in different settings. During the same time period, UK researchers began to explore differential effectiveness, finding empirical evidence of variation in schools' effectiveness with respect to different outcome measures (e.g. Mortimore *et. al.*, 1988; Smith & Tomlinson, 1989) and different subject areas (Fitz-Gibbon *et. al.*, 1989). Findings involving differential effects based on student factors such as ethnicity, gender, or prior attainment were inconsistent (Nuttall, Goldstein, Prosser, & Rasbash, 1989; Sammons *et. al.*, 1993); more recent developments in this specific area of EER are revisited in Section 3.1.2.

The 1980s also saw advances in multilevel modeling techniques. Earlier research had already begun to incorporate levels of aggregation into quantitative analyses (e.g. Alexander *et. al.,* 1981; Lau, 1979), and multilevel approaches were further improved through the work of Bryk and Raudenbush (1988, 1992) and others. These advances were supported by the creation of computer programs to support multilevel modeling by Bryk *et. al.* (1986) in the US and simultaneously by Longford (1986) in the UK (Teddlie & Reynolds, 2000). Multilevel modeling subsequently became a cornerstone of quantitative analysis in EER (Creemers, Kyriakides, & Sammons, 2010).

3.1.2 More recent developments in EER

EER has been repeatedly criticized for lacking theoretical foundations (e.g. Coe & Fitz-Gibbon, 1998; Thrupp, 2001). As Scheerens (1992) put it, much of the early research was characterized by "fishing expeditions for significant correlations" (p. 67).

A number of researchers in the 1990s attempted to address this criticism by creating conceptual models of school effectiveness (e.g. Creemers, 1994; Creemers & Reezigt, 1992; Scheerens, 1992; Teddlie & Stringfield, 1993). Sammons, Thomas, and Mortimore (1997), drawing on the work of Creemers (1994) and Scheerens (1992), also developed a theoretical model that included a departmental level, and emphasized department leadership as a critical link between school and teacher effectiveness.

The comprehensive model by Creemers (1994) has been particularly influential. This model incorporates factors with positive influence on student achievement at the school, classroom, and student levels, as well as student background characteristics, and accounted for the relationships between factors at the different levels (for example, school-level factors were supposed to have an indirect effect on student outcomes, mediated by classroom factors). The model outlines four formal principles of consistency, cohesion, constancy and control, although critics have suggested that consistency in particular might have been overemphasized where differentiation according to student groups and different learning tasks might be more appropriate (Jamieson & Wikeley, 2000).

The comprehensive model also informed other theoretical developments, notably the dynamic model conceived by Creemers and Kyriakides (2008). The dynamic model adds a fourth level, essentially the national or regional policy context, specifies dimensions for the measurement of factors at each level (frequency, focus, stage, quality, and differentiation), and addresses interactions between factors both within and across levels of the model. It also allows for a wider range of outcomes (cognitive and affective, for example), and incorporates development over time as a consideration at all levels. The influence of the dynamic model on approaches to school improvement research holds particular relevance for the

current study; links between school improvement and the EER knowledge base will be elaborated in more detail in Section 3.1.3.

The importance of these theoretical developments is that, in contrast to the wealth of earlier studies that investigated which and to what extent various factors were correlated with differences or predicted changes in student outcomes, these conceptual models take EER a step further towards exploring *why* factors at various levels were linked to differences in student outcomes. Empirical studies have sought to test the Comprehensive Model (e.g. De Jong, Westerhof, & Kruiter, 2004; Kyriakides, 2002) and the dynamic model, but as Scheerens (2013) points out in a recent review, there is still a "relative scarceness" (p.10) of systematic use of established theory in EER.

In addition to developing theoretical models, educational effectiveness researchers in the twenty-first century have had a vested interest in refining methodology to better address the complexities and dynamic nature of school processes and outcomes (Creemers, Kyriakides & Sammons, 2010).

An emphasis on equity, in terms of reducing gaps in outcomes and access to educational opportunities between groups of students, has played an important role in both theoretical and methodological advances. This concern for equity is not new; as early as the Coleman report (J. S. Coleman et al., 1966), there was evidence that school factors had a greater influence on disadvantaged or ethnic minority students, and similar conclusions have been supported by later research (Scheerens & Bosker, 1997). As Sammons (2010) notes (p.53),

> "There is increasing interest in the ways educational influences may compound existing inequalities related to student background, through unequal access to particular types of school, or differences in the quality of provision and the experience of teachers, for different groups of students."

Recent research has reflected this interest, investigating the relative attainment gap of Black Caribbean students as well as disproportionality in schools' identification of special educational needs across ethnic groups (Strand, 2010; Strand, 2011; Strand & Lindsay, 2009). Kyriakides and Creemers (2011) tested their dynamic model using equity criteria (the extent to which schools reduced the variance between students) and compared the results to what they had found earlier using

quality criteria (i.e. student achievement gains); not all of the same factors associated with quality were found to be associated with equity, and the authors advocate further research to identify additional factors specifically relevant to promoting equity. Kelly (2012) further asserts that the definition of equity itself requires further clarification, and suggests the use of additional measures (such as his Attainment Equity Index) to more thoroughly capture and interpret it.

Differential effects, with the potential to inform policy to support equal opportunities (Creemers, Kyriakides & Sammons, 2010), have been on the cutting edge of recent EER, and are central to the study of equity in education. Investigations of differential effects for groups of students were traditionally limited to one background variable at a time (Creemers, Kyriakides & Sammons, 2010). Recent empirical research, however, has considered possible interactions between different factors, and has shown evidence of significant interactions between such background factors as student ethnicity, gender, poverty, social class and prior attainment (Kyriakides, 2004; Strand, 1999, 2008, 2010). These findings suggest that a failure to consider interactions between factors risks misrepresentation, although presumably this applies to a greater or lesser extent depending on the substantive interest driving a particular analysis.

The development of approaches to account for single-level and cross-level interactions goes beyond differential effectiveness. Multilevel approaches to confirmatory factor analysis (B. Muthen, 1990) and structural equation modeling (SEM) (B. Muthen, 1994) have contributed to advances in this regard. Multilevel confirmatory factor analysis (MCFA) and multilevel SEM have enabled researchers to account for within-level as well as cross-level interactions, and for the construction of models with latent factors based on multiple indicators at multiple levels (e.g. Den Brok, Brekelmans, & Wubbels, 2006; Marsh et al., 2009; Toland & De Ayala, 2005). The use of these techniques, however, is somewhat limited by the necessity for an a priori theoretical framework. To address this, Van de Vijver and Poortinga (2002) built upon Muthen's approach to develop a related method of multilevel exploratory factor analysis (MEFA). Perhaps due to the complexities involved in interpretation, including the fact that subjective decisions play a role in determining results, MEFA has been used in relatively few empirical studies in education (D'Haenens, Van Damme, & Onghena, 2010).

The question of stability has also come into focus as a central consideration in EER (Kyriakides, 2004). While some studies have demonstrated fairly stable school effects over time (Gray *et. al.*, 1995; Raudenbush, 1989), others have shown evidence of variation from year to year, for example in departmental effects (Thomas, Sammons, Mortimore, & Smees, 1997). Stability over time is a key dimension of the dynamic model, and an area in which findings in EER have been inconsistent to date (Creemers & Kyriakides, 2008). Considerations of stability have led longitudinal models to gain prominence in EER. Another reason for the use of longitudinal models relates to the importance of measuring longer-term effects, as demonstrated by research such as the study of primary schools effects on secondary school outcomes by Goldstein & Sammons (1997).

Consistency, understood here as effectiveness across different types of outcomes, has also been identified as an area for growth in EER (Kyriakides, 2004). In Section 3.1.1, early studies providing evidence of differences in school effects across different academic departments were mentioned. Beyond considerations of differences across subject areas, however, EER has tended to rely heavily on one type of outcome, achievement, with relatively few exceptions (e.g. Kyriakides, 2005; Melhuish et al., 2008). Further research is needed, not only to explore different kinds of outcomes other than academic achievement, but also to gain insight into possible relationships between different types of outcomes (Creemers, Kyriakides & Sammons, 2010).

Despite advances in theory and methodology, criticisms of EER continue. Some critics have suggested that EER does not adequately address issues of social justice and political context (e.g. Wrigley, 2013), despite the strong equity focus in EER that has been noted by a number of authors (e.g. Sammons, 2010). Others question EER methodology; a recent example is Gorard's (2010) critique of school effectiveness, which seemed to be based on a lack of understanding of methodology, measurement error, and a lack of awareness of the literature (Muijs, Kelly, Sammons, Reynolds, & Chapman, 2011).

3.1.3 School improvement

School improvement emerged out of a completely different paradigm from EER, with closer ties to implementation and practice than academic inquiry and studied using predominantly qualitative approaches.

Although there were school improvement efforts in the 1960s and 1970s, these consisted mainly of top-down curriculum reform (Reynolds, 2004). The first efforts to connect evidence from EER to school improvement programs in the US were based on the findings of Edmonds (1979) regarding characteristics of effective schools, which were translated into a variety of programs for school improvement (e.g. Clark & McCarthy, 1983; McCormack-Larkin & Kritek, 1982). These efforts were still subject to the criticism that the underlying research had an inherent sampling bias. The later introduction of context factors in EER to address this issue led to context-sensitive differentiation of school improvement programs as well (Teddlie & Stringfield, 2006).

The use of research evidence in school improvement in the US shifted into the phase of school restructuring in the 1990s, presented as an alternative to models of improvement rooted in effective schools research (Murphy, 1992). Restructuring emphasized site-based management, changes in the structure of teaching, parental involvement, and alternative approaches to accountability and scheduling (Ibid). Critics of school restructuring asserted that these reforms failed to have a meaningful impact on learning (Teddlie & Stringfield, 2006).

While the use of evidence to design school improvement initiatives began early in the development of EER, the approaches to school improvement research (SIR) continued to rely on qualitative case study methods until recently. When researchers did use quantitative methods to assess the effects of improvement programs on academic outcomes (e.g. Borman, Hewes, Overman, & Brown, 2003; Slavin & Madden, 2001), they were criticized for failing to account for processes. The use of mixed methods has been proposed by some researchers to address these issues (e.g. Sondergeld & Koskey, 2011), but mixed methods research presents its own set of methodological challenges (Creswell & Plano Clark, 2011). The relationship between EER and SIR, in short, has been fundamentally problematic. The importance of striving for better correspondence between the two has been

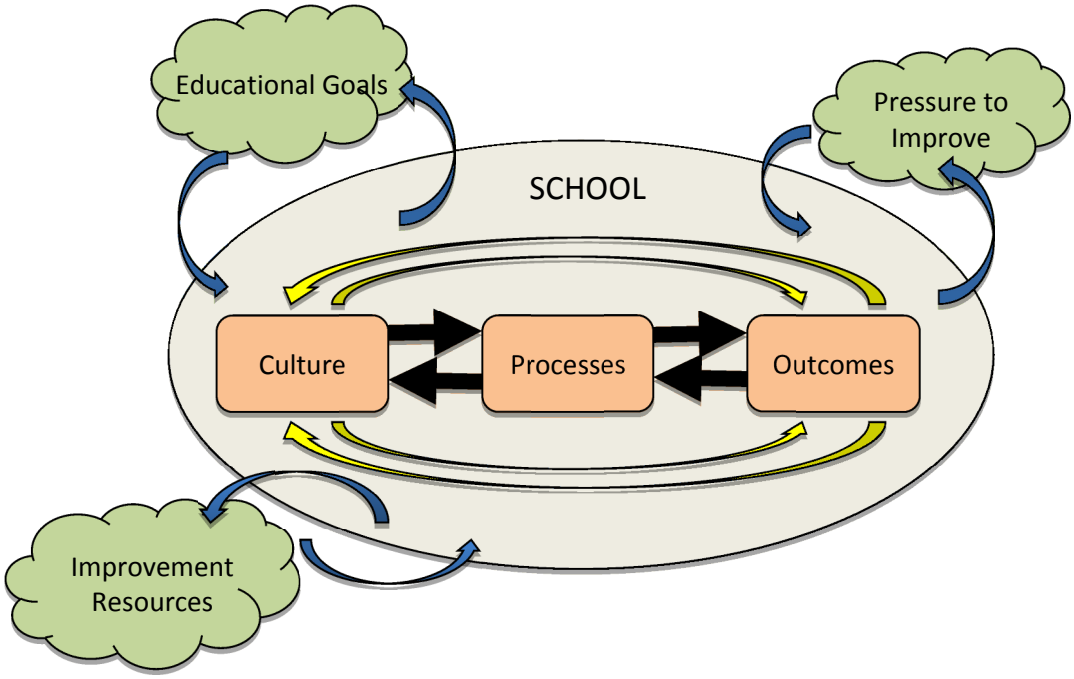
highlighted as an important direction for emerging research (Creemers, Kyriakides & Sammons, 2010; Reezigt & Creemers, 2005; Reynolds & Stoll, 1996; Teddlie & Reynolds, 2000).

The Dynamic Approach to School Improvement (DASI) (Creemers & Kyriakides, 2012), which incorporates steps for the implementation of improvement projects adapted from the dynamic model of educational effectiveness, represents one of the most thorough attempts to develop conceptual connections between EER and school improvement to date. This approach to improvement highlights the importance of an emphasis on student learning as the ultimate goal of improvement, the necessity of school self-evaluation, and the involvement of stakeholders in designing improvement plans based on knowledge of relevant factors from the dynamic model (Creemers, Kyriakides, & Antoniou, 2013). DASI benefits from its intrinsic link to a comprehensive theoretical framework with empirical support (e.g. Creemers & Kyriakides, 2010; Kyriakides & Creemers, 2008), which allows for systematic evaluation of improvement drawing on the dynamic model of effectiveness. DASI has been used in studies of school improvement (Antoniou & Kyriakides, 2011; Demetriou & Kyriakides, 2012), but further research on this approach in a variety of settings would be needed to test the authors' claims that DASI is flexible enough to address the different needs of individual schools in varying contexts. The relevance of this framework to the current study (which was originally considered) would have been contingent upon findings indicating a high degree of involvement of various stakeholders in improvement planning and evaluation of progress toward improvement goals; before embarking upon qualitative inquiry, this still represented an unclear aspect of CFN participation and purpose.

The Effective School Improvement (ESI) framework (Reezigt & Creemers, 2005), based on the Effective School Improvement Project (ESIP) conducted across eight countries (Reezigt, 2001), offers another synthesis of EER and school improvement into a framework that can be adapted for use in the design and implementation as well as the empirical study of school improvement. The ESI framework is intended to be flexible enough for use in a variety of settings, so that the only prescriptive feature is that student outcomes or factors with clear influence on student outcomes, however they may be defined, must be the ultimate goal of

improvement. Figure 1 below, adapted from Reezigt and Creemers (2005), illustrates the various elements of the ESI framework.

Figure 1 Major components of the ESI framework



Three contextual factors are included in the framework: pressure to improve (further separated into factors of market mechanisms, external evaluation/accountability, external agents, and the participation of society in educational change), resources (collectively defined by factors of school autonomy, local support, and financial resources and favourable working conditions), and formal educational goals. At the school level, the framework draws from previous research including reports from the International School Improvement Project (ISIP) (D. Hopkins, 1987) to define three factors: improvement culture (defined by features of shared vision, autonomy used by schools, ownership of improvement, collegial collaboration, leadership, staff stability, time for change, improvement history, and internal pressure to improve), improvement processes (defined by assessment and diagnosis of improvement needs, planning and implementation, evaluation and reflection), and improvement outcomes (defined by changes in the quality of school, teacher, and student outcomes). Explicit consideration of context, together with the specification of factors at multiple levels without limiting the criteria that can be used to measure them, are particular strengths of the ESI framework. This can be seen as an appropriate theoretical foundation for

researching improvement in a network of schools, with notions of improvement cultures, processes and outcomes extended to apply to networks as well as to schools.

3.2 LEADERSHIP AND LINKS TO EER

The influence of leadership has been widely documented in EER as well as in studies of school networks. As such, leadership plays a crucial role in the current study. The following section provides a brief review of key developments in theory pertaining to educational leadership and an abbreviated analysis of recent empirical research, as well as extensions of leadership theory from individual schools to networks.

3.2.1 Dominant leadership theories

There are numerous definitions for the term "leadership", and it is often conflated with management or power. For the purposes of this discussion, leadership can be broadly understood as "the process of influencing others to understand what needs to be done and how to do it, and the process of facilitating individual and collective efforts to accomplish shared goals" (Yukl, 2006, p.8). Note that this definition does not limit leadership to a single individual, nor does it restrict the notion to formal job roles.

Three main constructs have dominated the literature in educational leadership: instructional, transformational and distributed leadership (Hoy & Miskel, 2013). Instructional leadership theory, which was originally framed mainly as individual leadership practice, emerged out of EER and associations between strong leadership and student outcomes (e.g. Edmonds, 1979; Leithwood & Montgomery, 1986). This view of leadership focuses on support for teaching and learning, by way of defining a school's mission, managing the instructional program, and promoting a positive learning climate (Hallinger, 2003). A review by Robinson (2010) identifies instructional leadership capabilities generally as content knowledge, building relational trust, and problem solving, but demonstrates some of the difficulties involved in specifying and measuring these capabilities. Critics of instructional leadership (e.g. Dimmock, 1995; Fullan, 1992) have dismissed it as both prescriptive and ill-matched to the realities and needs of schools. However,

there has been considerable evidence that the impact of instructional leadership on outcomes such as student outcomes may be greater than that of other types of leadership (e.g. Robinson, Lloyd, & Rowe, 2008). Instructional leadership has also been more broadly defined in recent research to include the notion of shared leadership (Louis, Dretzke, & Wahlstrom, 2010).

Transformational leadership is a concept borrowed from the work of Burns (1978) and Bass *et. al.* (1987) and adapted to educational contexts by Leithwood and Louis (1998). While incorporating many of the elements of instructional leadership, the transformational perspective includes additional dimensions. Leithwood and Jantzi (2006) propose a framework for transformational leadership based on three categories of practice: setting direction, developing people, and redesigning the organization. Additionally, this model does not specify leadership as an individual role, rather, it allows for leadership to be shared, which shows an overlap with distributed leadership (Leithwood & Jantzi, 2000; Leithwood, Louis, Anderson, & Wahlstrom, 2004; Marks & Louis, 1997).

Distributed leadership has been conceptualized in a variety of different ways. Spillane, Halverson, and Diamond (2004) defined distributed leadership as interactions between leaders, followers, and situations. Gronn (2002) posited, based on a review of selected literature, that distributed leadership may be enacted through spontaneous collaboration, intuitive working relations, and institutionalized practices. Leithwood, Harris, and Hopkins (2008) further drew on a review of empirical literature in education to support claims that a wider distribution of leadership has more impact on student outcomes, and that some patterns of distribution are more effective than others.

Empirical studies of transformational and distributive leadership have shown evidence of relationships to student learning. Many have drawn on the EER knowledge base, employing multilevel and longitudinal models (e.g. Hallinger & Heck, 2010) and mixed-methods approaches (Day et al., 2011; Leithwood et al., 2009; Mulford & Silins, 2011), incorporating factors of school effectiveness identified in prior EER as mediators of leadership effects, and in some cases attending to outcomes other than achievement such as engagement and motivation (Day *et al.*, 2011).

Hallinger (2003) notes that the measurement of leadership that is not restricted to an individual is inherently challenging. Although some studies have been limited by an overreliance on teacher perceptions to measure leadership variables (e.g. Leithwood & Jantzi, 2006), others have addressed measurement challenges by including pupil, principal and/or parent perceptual measures as well (e.g. Hallinger & Heck, 2010; Sammons, Gu, Day, & Ko, 2011; Leithwood *et al*., 2009). There has been a general consensus across these studies that the effects of distributed and transformational leadership on student outcomes are indirect and relatively small but significant.

Hallinger and Heck (2010) found that the effects of leadership are larger on changes in student achievement than student achievement at a single point in time. There has also been empirical evidence of stronger leadership effects on school processes such as organizational learning (Mulford & Silins, 2003) and teacher motivation, capacity and work setting (Leithwood & Jantzi, 2006) than on student outcomes. The work of Sammons *et al.* (2011) provides support for the study of leadership strategies by finding the effects of specific dimensions of leadership on pupil outcomes via mediating factors such as improvements in school conditions and teacher collaborative culture. All of these findings have relevant implications for the current study, as they highlight the significance of leadership in SIR.

Various authors have extended theories of transformational and distributed leadership beyond individual schools. Fullan (2006) emphasizes "strong connections and relationships with other parts of the system" (p.96) as an important feature of effective school leadership. School leaders often take on complex and multiple roles in networking and collaboration, and as a result it makes sense that studies of networks often have a strong focus on the nature of those roles or on the attitudes and experiences of these individuals (e.g. Coleman, 2012; Hopkins & Higham, 2007; Lindsay et al., 2007). Higham, Hopkins, and Matthews (2009) outline three dimensions of leadership necessary for networking and collaboration, specifically system leadership (involving leading and contributing to innovation beyond an individual school), distributed leadership in and across schools, and leadership of learning (i.e. prioritization of teaching and learning as the focus of improvement). Distributed leadership as an important feature of successful networks has been emphasized by multiple researchers (Harris, 2006; Jackson,

2006). Interestingly, however, there have also been indications that firm hierarchical leadership is needed in the initial formation of a network, after which a shift to distributed leadership may be more appropriate (Chapman & Harris, 2004; Hadfield, 2007).

3.3 SCHOOL NETWORKING AND COLLABORATION

Having addressed definitions of school networks in Section 1.1.1, the following two sections present a discussion of theory in school network research and a critical analysis of studies of school networks.

3.3.1 The place of theory in researching networks

Theories from outside of education may be useful to inform the study of school networks. For example, constructivist theory, based on the ideas of Vygotsky about cooperative learning (Vygotsky, Vygotsky, & John-Steiner, 1978), implies an emphasis on how collaboration contributes to professional learning beyond what would occur in individual schools, while social capital theory (Lin, 1999) suggests a focus on the sharing or transfer of resources and the benefits to individual schools or whole networks as a result of collaboration. By contrast, a functionalist perspective (Munch, 2001) might emphasize the factors and configurations that make networks more or less effective at achieving the purposes for which they were established.

Despite the availability of theories relevant to networking in general, the study of networks in education has until recently been "largely a-theoretical and evaluation based" (Muijs et al., 2010). Two approaches to emergent theory specific to school networks are presented below.

In the context of Networked Learning Communities (NLC) in England, Katz and Earl (2010) proposed a 'theory of action' based on the principle that changing student learning depends on changing the practices and structures in schools, which in turn depends on professional learning through interaction within and between schools (Ibid). Within this framework, the success of a network is seen to depend on purpose and focus, relationships, collaboration, enquiry, leadership, and capacity building and support. This theory has not been widely applied beyond the scope of NLC's, and might benefit from further development and testing in other

settings, particularly considering that research has shown little evidence of impact of NLC's on student outcomes (Sammons, Mujtaba, et al., 2007).

Providing what might serve as a point of departure for broadly applicable theoretical frameworks for networking and collaboration in education, Muijs, Ainscow, Chapman, and West (2011) drew upon theory in sociology and psychology as well as previous empirical research to create a comprehensive network typology with the following dimensions:

- Goals and activities
- Voluntarism or coercion
- Power relations
- Network density
- External Involvement (often by universities or charities)
- Time frame
- Geographical spread
- Density of schools
- Within- and between-school networking
- Diffuseness
- Formalization

This conceptual framework was used in the current study to guide a descriptive as well as thematic analysis of network structures, roles, participation and purpose.

3.3.2 Studies of networks

Despite the ubiquity of school networks and claims about the promise of collaborative structures to support organizational learning (OECD, 2000), networks vary so widely in their structures, contexts, and goals that there is considerable inconsistency in the research literature on the impact of networks on teaching and learning (Bell et al., 2006).

In terms of network influence on practitioners, there has been some evidence that collaboration between schools can improve the knowledge and skills (Adler, Cragin, & Searls, 1995; Bielefeldt, Moursund, Underwood, & Underwood, 1999; Gettinger, Stoiber, & Lange, 1999; Riley & Jordan, 2004) as well as attitudes and

motivation (Peters, 2002; Pinon, Samii-Shore, & Batchelder, 2002; Thurlow, 1999), of teachers and principals. Further, some previous research has found that networks can facilitate implementation of new initiatives by providing a 'hub' of information and resources (Wohlstetter et al., 2015). However, other studies have suggested at best inconsistent impact on classroom practice, particularly in networks with externally imposed goals (Jolliffe & Hutchinson, 2007; Kahne, O'Brien, Brown, & Quinn, 2001; Sanders, 1999), and even that networking can have a negative effect on teachers by imposing what they perceive to be an extra burden of responsibility (Wohlstetter et al., 2003).

Similar variation exists in findings regarding student outcomes, ranging from indications that networking is associated with gains in attainment (e.g. Adler, 1995; Zetlin, MacLeod, & Michener, 1998) and affective and social outcomes (e.g. Howley-Rowe, 2000; Laferriere *et al*., 2010), to evidence that network initiatives failed to connect with student learning (Kahne *et al.*, 2001). Much of the evidence base has relied on teacher perceptions of improvement in student outcomes, though some small-scale studies have used more direct measures (Greenberg, 1996; Montgomery, 2001). Large-scale studies of NLCs (Katz & Earl, 2010; Sammons, Mujtaba, et al., 2007) and Federations (Lindsay *et al*., 2007) in the UK using student attainment data have generally indicated not much substantial variation within networks (Sammons et al., 2007)and little evidence that attainment gains are higher in networked schools. There is a dearth of research in this substantive area using multilevel modelling approaches to investigate network effects, with a few exceptions (e.g. Chapman & Muijs, 2013, 2014; Fruchter et al, 2015). The latter study, notably, focused on CFNs in New York City, but only took into account the school and network levels of the school system in multilevel models intended to analyse the association between networks and student achievement. Although in this particular case the researchers still came to the conclusion that there was little effect of networks on student outcomes, this approach constitutes a misspecification of the underlying multilevel structure by omitting the student level, which risks obtaining larger variance estimates at higher levels as a result of aggregation (and in turn risks reporting faulty evidence of effects on student achievement).

The vast majority of school network studies have relied on qualitative case study methodology (e.g. Chapman, 2008; Wohlstetter et al., 2003). While findings from these may be meaningful, they are extremely context-sensitive, and typically driven by purposive sampling of relatively successful networks. Thus, there have been extremely few studies of 'bad' or 'weak' networks (Black-Hawkins, McIntyre, & McLaughlin, 2004). Others have additionally made use of large-scale surveys of various stakeholders (e.g. Kahne et al., 2001; Katz & Earl, 2010). This approach is limited by an overreliance on the perceptions of respondents, which may differ from actual practice and results, and may also depend heavily upon which stakeholder groups' views are included.

In short, there is certainly a need for more research to explore the connections between networking for school improvement and changes in student outcomes, particularly beyond the UK context.

3.4 CONCLUDING REMARKS: TOWARD A THEORETICAL PERSPECTIVE

A major contribution of the present study is that it applied rigorous multilevel modelling techniques drawn from the EER knowledge base to a study of school networks. It incorporated both quantitative and qualitative methods to explore the phenomenon of school networking and collaboration, drawing on a perspective rooted in both educational effectiveness and school improvement theory. This blending of theoretical perspectives has ample precedent in these two increasingly overlapping fields, and combining both perspectives has the potential to contribute deeper understandings to a still-emergent area of education research (networks and collaboration). By taking an exploratory approach, this study allowed for the possibility that connections, collaboration, and leadership might exist in networks beyond the structures mandated by policy. This type of potential complexity led to a flexible approach to the initial application of theory – the ESI framework informed the selection of topics for interview schedules, for example, but without basing this on a tacit assumption that network participation, activity and outcome would fall neatly into line with this framework, as it was not initially known whether or not CFNs in fact functioned as networks or as school improvement mechanisms.

The specific elements of the ESI framework did, however, form a basis for the theoretical conceptualisation of this research. The quantitative strand focused on student achievement outcomes based on the emphasis on state assessment performance in accountability in the context for the study; this aligned with both the focus on student outcomes as the focus of school improvement in the ESI framework, and the claim that CFNs could help schools to raise achievement that formed the basis for one of the main aims of this research. Because these same student achievement outcomes played such a prominent role in NYCDOE accountability, and were used in the periodic evaluations of both CFNs and schools, the study conceptualised these as both overarching *educational goals* at the system level and *outcomes* at the school level. However, to allow for the possibility that stakeholders within schools and networks conceived of educational goals in broader or different ways, interviews were designed to elicit open-ended feedback from these stakeholders regarding their own understandings of educational goals and priorities for improvement. *Processes* and *culture*, both at the school and network levels, constituted foci for qualitative investigation, as previous research on both school improvement and inter-school collaboration have indicated that the cultures within and between schools, and the specific structures and types of interaction between actors in a network of schools, are crucial in determining the extent to which collaboration takes place and the impact it has on schools. Because the focus for this research was on the network structure as a whole, and not limited to individual schools, the contextual factors from the ESI framework were especially fundamental considerations in the design of the study. Qualitative enquiry included the exploration of external *pressure to improve* and the ways in which this affected participation in networks and perceptions of their impact, as well as the extent to which CFNs functioned as or provided *improvement resources* and whether schools were seeking or accessing additional improvement resources not associated with CFNs. This mapping of the ESI framework onto the conceptualisation of this study suggests that the framework was, in a sense, extended to have an additional outer ring (if visualised according to the diagram in Figure 1, p.28), so that the same factors at the school level were again investigated with regard to the CFN level.

Having addressed the theory framing this research and the knowledge base to which it sought to contribute, it is also important to address the limitations of this study as they relate to previous research. The variables considered in the quantitative analysis were limited by the availability of secondary data, precluding the inclusion of a class or teacher level in the quantitative models used, or of student outcomes other than achievement, despite the fact that these have been identified as important considerations in EER. Additionally, any causal claims about the relationship between networks and school improvement must be at best tentative and cautiously interpreted. It would be inappropriate to seek to make strong causal claims about networking and school improvement, because schools are affected by a "multiplicity of different influences" for which it is impossible to control absolutely (Sammons et al., 2007). This applies also to any generalizations beyond the specific context based on either or both quantitative and qualitative findings. Just as the theoretical perspective for the study was not a prescriptive one, the contribution of the study to theory will provide a point of departure for dialogue and further inquiry about networks but not a definitive set of answers about cultures, processes and outcomes of school networks beyond the immediate context.

4 METHODS

4.1 CHAPTER GUIDE

This chapter is intended to:

- Introduce the aims and research questions guiding the present study
- Present the overall research design and approach, linking it to the aims and research questions
- Describe and justify the specific data sources, methods, and analytical approaches used
- Explain the types of and approaches to mixed methods integration at various stages of the study
- Discuss researcher positioning, epistemological stance and the theoretical perspectives
- Comment on ethical considerations and access issues
- Address the limitations of the study

4.2 AIMS AND RESEARCH QUESTIONS

The two overarching aims of this study are to address the claims that Children First Networks (CFNs) facilitate collaboration between schools and that they help to raise student achievement. To this end, this research was designed to investigate possible relationships between CFNs and student achievement, and to understand the processes and interactions that operate within these networks and the extent to which they play a role in promoting school improvement. The study is guided by the following research questions:

RQ1: Is there evidence to suggest that networks have measurable effects on student and school achievement and progress?

- **RQ1a:** How does 8th grade achievement in literacy and mathematics vary between networks, districts, schools, and students, and is there evidence of significant overall effects at these levels?
- **RQ1b:** To what extent can variation in 8th grade achievement be statistically explained at each level by student background factors, prior attainment in 5th grade,

school context and composition, and network or network-affiliation characteristics (e.g. duration of affiliation)?

- **RQ1c**: How consistent are apparent effects within years across subjects (literacy and mathematics), and how stable are they within subjects across years?
- **RQ1d:** How do answers to the above questions compare when considering changes (at the school level) in achievement over time, rather than achievement status in one year?

RQ2: How are the roles and functions of networks understood by various stakeholders (network staff, principals and teachers), and what are the processes and interactions that operate within networks and the network structure as a whole?

- **RQa:** How do various stakeholders (network staff, principals, teachers) describe network activities and support offerings, and to what extent do these descriptions align across stakeholder groups?
- **RQ2b:** To what extent, in what ways, and with whose participation does collaboration and interaction happen within networks?
- **RQ2c:** What processes and which stakeholders drive network affiliation and shape network support for schools?
- **RQ2d:** What are the perceptions of various stakeholders about the impact of networks on teaching, learning (e.g. student achievement outcomes), and/or leadership in schools?

RQ3: How do understandings of processes, interactions, values and priorities within schools and networks help to explain relationships (or lack thereof) between network affiliation and student achievement?

4.3 MIXED METHODS RESEARCH DESIGN

In order to address the aims and questions introduced in the previous section, the research design is conceived as a mixed methods (MM) study, combining quantitative and qualitative strands of inquiry.

4.3.1 Why mixed methods?

Johnson, Onwuegbuzie, and Turner (2007) define MM research as "an intellectual and practical synthesis based on qualitative and quantitative research…that often will provide

the most informative, complete, balanced, and useful research results" (p.129). This aptly summarizes the basic rationale for the use of mixed methods in the present study.

Pragmatism, an epistemological stance traditionally associated with MM research (Teddlie & Tashakkori, 2003), rejects the notion of antagonism between quantitative and qualitative "paradigms" and their associated discourses (Morgan, 2007), avoiding limiting definitions of "reality" or "truth" and focusing instead on the best approach to get results (Creswell & Plano Clark, 2011). A pragmatist stance informs the current research design, defined by an emphasis on the choice of methods best suited to addressing the research questions. In the context of this research, prioritizing the pragmatist "dictatorship of the research question" (Tashakkori & Teddlie, 1998; 2003) is viewed as a realistic and useful way of reconciling methods to the challenge of addressing qualitatively-oriented questions focusing on understanding and illuminating processes and perspectives as well as quantitatively-oriented questions focusing on outcomes and effects. In short, quantitative and qualitative methods are selected based on their fit with the questions being asked, not based on broader assumptions about the nature of knowledge or understanding linked specifically to quantitative or qualitative approaches. The implications of this are considered in more detail in Section 4.7, in which additional considerations of epistemological positioning and theoretical perspective are discussed in greater depth.

4.3.2 Classifying the research design

Various typologies have been developed to characterize and classify mixed methods research with differing emphases and degrees of specificity (e.g. Creswell et al., 2003; Creswell, 2008; Greene, Caracelli, & Graham, 1989; Greene & Caracelli, 1997; Leech & Onwuegbuzie, 2009; Tashakkori & Teddlie, 1998, 2003). A particularly well-known and frequently-cited approach is to categorize MM research designs according to four dimensions: the level(s) of interaction between strands, their relative priority, timing, and mixing procedures (Creswell & Plano Clark, 2011). According to these dimensions, the present study may be described as a variation on the sequential explanatory MM design (Creswell et al., 2003). The quantitative strand is prioritized based on the prominence of assessment outcomes in accountability measures in the New York City context, and strands are approached sequentially rather than concurrently. Mixing happens in multiple ways; quantitative findings inform sampling for the qualitative strand, which yields insight

into underlying processes and contexts that help to explain – though not in a statistical sense – and enhance understanding of the quantitative findings.

Based on the work of Greene, Caracelli and Graham (1989), MM research can also be classified according to purpose. Within their framework and according to their definitions, another way of understanding the choice of design for this study is that it aims primarily for complementarity, in which "overlapping, but also different facets of a phenomenon" are considered (p.258), and expansion, which "extends the breadth and range of the study" (p.259). To a lesser extent, other purposes described by Greene et al. also apply (triangulation, initiation, and development), but are less central to study's aims.

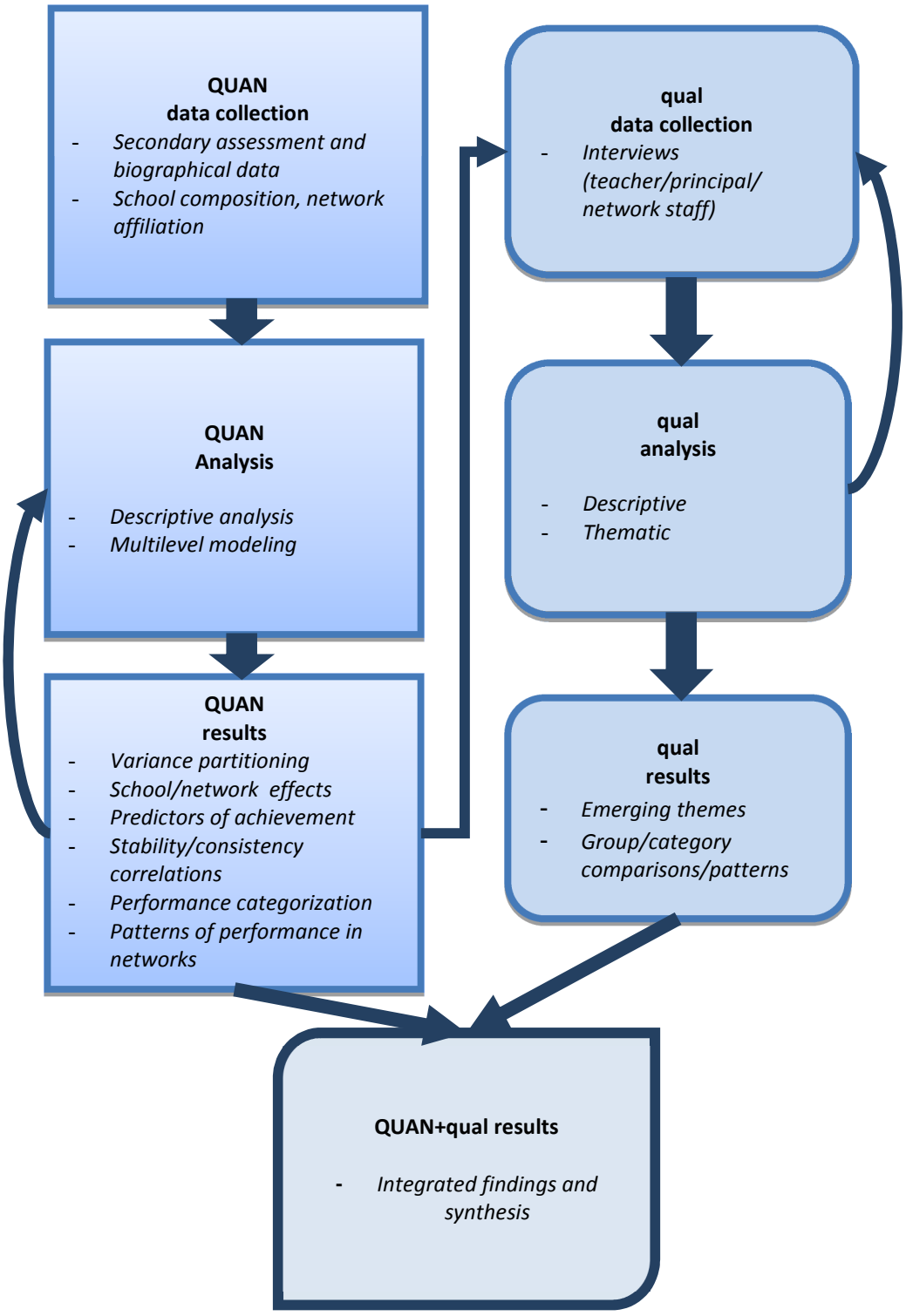
The language of these typologies is useful to help the reader to gain an immediate grasp of the type of design and purpose of this study. However, it is important to recognize that these classifications are perhaps more clear-cut than the complexities and soft boundaries of the actual research process. Whereas in a strictly-defined sequential explanatory design, quantitative and qualitative strands of the research would constitute distinct phases, in practice, quantitative and qualitative strands were not strictly sequential in this research. There was some temporal overlap, particularly between quantitative analysis and qualitative data collection and analysis, and there was an iterative aspect to the analytical process in that findings from one strand often suggested ideas for improving or expanding analyses for the other. This type of "dialectical" approach has unique potential to produce what are referred to as "synergistic understandings" (Sammons, 2010). For research in educational effectiveness and school improvement, this process has the potential to facilitate broader and deeper interpretations of effectiveness and improvement in particular contexts (Teddlie & Sammons, 2010).

4.3.3 Summary of the specific design

A brief overview of the research design is provided in the procedural diagram in Figure 2, illustrating how (and when) quantitative and qualitative strands are integrated. Conventions developed by Morse (1991; 2003) are followed, including abbreviations for "quantitative" and "qualitative" (QUAN and QUAL), use of uppercase or lowercase to signal relative priority of strands, "+" signs to indicate simultaneity, and arrows to indicate sequential procedures. Additionally, the different shapes used to represent quantitative and qualitative research activities and products are a variation on the conventions used by mixed methods researchers for such diagrams (see Tashakkori & Teddlie, 1998).

The first stage of the research project after the planning process (beginning at the top left of the diagram) was to obtain and analyze secondary data, which fed into the sampling design for the second phase, which was subsequently approached using qualitative case study methodology. Arrows are used to identify points at which the two strands were integrated, and also to signify that several phases of the research were not strictly sequential (e.g. qualitative data collection and analysis were undertaken iteratively).

Figure 2 Procedural diagram



4.3.4 Remarks on integration

Integration of quantitative and qualitative approaches is a key aspect of any MM design. While achieving meaningful integration may be methodologically and practically challenging, and some authors have expressed concerns about the extent to which mixed

methods researchers actually integrate their results (Bryman, 2007; Yin, 2006), increasing numbers of exemplars of integration and its benefits are apparent in the existing literature. In educational effectiveness and school improvement research, the opportunities afforded through integration for broader interpretations and more complete understandings, including relationships between processes and outcomes, have been illustrated by recent empirical studies (Day et al., 2008; Kington et al., 2011; Sammons et al., 2014; Sammons et al., 2007; Sondergeld & Koskey, 2011).

Integration may take place at one or more stages of the research process, using a wide variety of strategies. Sequential explanatory designs are typically characterized by a great deal of separation between strands; data collection and analysis take place in separate phases, and results are often reported separately and integrated only in a final discussion section (Creswell et al., 2008). In the present study, an effort was made to achieve integration at several stages of the research process, including design, data collection, analysis, and interpretation. At the design stage, the reader may note that while there are separate blocks of questions that may be seen as implicitly qualitative or quantitative, there is also a specifically mixed question (RQ3). This helped to set the stage for thinking about the project within a MM framework from the outset. At the data collection stage, as illustrated in the procedural diagram, integration took place when quantitative findings are used to inform purposive sampling for the qualitative strand. At the analysis stage, integration took several forms, including but not limited to some qualitizing of quantitative data and quantizing of qualitative data.[5] As noted by Bazeley (2012), combining and converting between data sources and types allows for "blended" results that maximize the potential for useful conclusions (this is referred to as "cross-over analysis" by Onwuegbuzie, Johnson, and Collins (2009)). Finally, at the interpretation stage, integration is achieved through discussion of how the qualitative findings explain and elaborate upon some of the quantitative findings, and implications of findings from both strands for areas of further inquiry. The goal of these efforts to combine and synthesize approaches is ultimately to facilitate synergistic understandings, which require not only a technical integration of methods but also conceptual integration arising from a dialogic interplay between strands (Day, Sammons, & Gu, 2008).

[5] "Quantizing" and "qualitizing" are terms in the mixed methods research tradition referring to transformations between types of data (Hesse-Biber, 2010).

Given this overview of the research design as a whole, the sections that follow elaborate upon the specific data, samples, and methods for qualitative and qualitative strands of the study.

4.4 THE ROLE OF THEORY ACROSS STRANDS

Quantitative inquiry in this study takes an educational effectiveness perspective. The approach is largely based on a balance struck between best practices established by prior research and limitations of the available data. The place of theory in this line of inquiry is in driving appropriate methodological choices, defining specific variables to consider in analyses, and contextualizing results within the existing literature pertaining to educational effectiveness.

The dynamic model of educational effectiveness proposed by Creemers and Kyriakides (2008) provides a particularly well-established conceptual and theoretical foundation for educational effectiveness studies, and has been tested in multiple settings (see Chapter 3 for further detail). It is not possible for this study to implement many aspects of the dynamic model in practice because, for example, while the model emphasizes the importance of the classroom level, the statistical data available do not include information at this level; it only includes the student, school, district and network levels. Nonetheless, the dynamic model provides a frame of reference for understanding how different levels, dimensions and features of schools and their broader contexts may affect student outcomes, and for this reason is used as a point of departure for the quantitative strand of this study, and a basis for addressing limitations.

The qualitative approach and methods draw on substantive theories from a considerably different knowledge base, but the role of theory is analogous to that described above for quantitative inquiry (as a frame of reference or as a basis for comparison and interpretation). Instead of defining variables, established theories relevant to school networks and collaboration and networking in education provided some a priori themes which are used, to a limited extent, to code interview and documentary data. Themes emerging from the data in a more grounded approach to analysis are examined against the backdrop of broader relevant theory, and compared across units of analysis (e.g. groups of stakeholders, individuals, or schools). This process has the potential to uncover patterns, exemplars, anomalies, or complexities indicating a need for further theoretical development. Theory was also essential to the process of developing research instruments;

the questions and topics in the interview guide, for example, are largely driven by existing networking and collaboration theories. The typology of networks provided by Muijs et al. (2011), which is mainly based on UK research but broadly-framed enough to be useful in other contexts, provides a framework for thinking about networks and characterizing them in terms of their purpose , structure, and dynamics (Section 3.3.1. As Merriam (1988) asserts, theory (understood both as substantive theory and as theoretical perspective) frames every step of the process in case study methodology, from design and choice of methods to analysis and interpretations.

A discussion of theory in a more general sense, i.e. theoretical perspective, is addressed in Section 4.7.1.

4.5 QUANTITATIVE DATA, SAMPLES AND METHODS

4.5.1 Secondary data

As indicated in the previous section, the first, quantitative strand of the study involves statistical analysis of secondary data. School-level data are publicly available on the NYCDOE website, while student-level data is considered confidential and had to be requested formally from the NYCDOE's Research and Policy Group (RPG) along with documentation to clarify interpretation of included variables and values. Data relevant to schools' network and cluster affiliations are not overtly classified by the NYCDOE as sensitive or confidential, but were obtained by special request to the RPG as they were not published online.

The first data request covered six academic years from 2006-7 to 2011-12, inclusive. These cutoffs were determined based on a number of practical considerations, including:

- Availability at the time of the request (assessment data is available only after the fall semester of the academic year following the administration of a particular test),
- A need for appropriate prior attainment measures to investigate student progress (i.e. 5^{th} grade scores were used from three years prior to each 8^{th} grade assessment year considered)
- Timing of the phenomenon being studied (the network structure that forms the focus of this research has only been fully implemented since 2010-11, so that 2009-10 constituted a sort of baseline year).

The data request was later expanded to include data from 2012-13 when the relevant assessment results were released. This was done for two reasons: First, to ensure that results were as current as possible at the time of writing up, and second, to take into consideration some of the relevant policy changes taking place while this research was being conducted (e.g. the rollout of the Common Core State Standards and associated changes to assessments as noted in the Context chapter, Chapter 2).

The following includes a brief description of each data source. Table 1 classifies these sources by the ways in which they were obtained.

- ***Biographical records (student)***: A separate dataset for each academic year includes, for each student, a scrambled student identifier, school of enrolment as of June 30th of the appropriate year, grade level, gender, ethnicity, home language, birthplace, English proficiency status, disability, special education program, free or reduced lunch eligibility, attendance (counts of days present and absent for fall and spring semesters), birth year and birth month.
- ***Assessment records (student)***: A separate dataset was provided for each academic year, containing each students' scrambled student identifier (matched to biographical records) along with variables relevant to literacy and mathematics assessments including test form, test grade level, testing accommodation(s) indicators, raw score, scale score, proficiency level, and performance rating.[6] In addition, each record included information related to biographical records, including the identifier(s) of the school(s) at which each student was recorded as having taken state literacy and mathematics assessments, grade level, gender, ethnicity, English language proficiency status, disability, testing modifications for literacy and mathematics, free or reduced lunch eligibility, home language, and attendance (days present and absent for the relevant year).
- ***Assessment performance (school, district and citywide)***: In New York State, students in grades 3-8 take annual standardized assessments in literacy and mathematics (other subjects are also assessed, but these two are

[6] Proficiency levels from 1-4 (1 being the lowest) are assigned according to cut scores each year. While these are used for school accountability in context, they are not comparable from year to year according to the technical reports for the relevant assessments (e.g. Pearson, 2012a).

prioritized for accountability purposes by city and state authorities). Data downloaded from the NYCDOE website includes variables for mean scale score, total students tested, and number and percent of students earning each proficiency level (1, 2, 3, 4) for each grade level in each assessment year (2007 to 2012), aggregated at school, district and citywide levels.

- ***Demographic snapshots (school, district and citywide)***: A workbook downloaded from the NYCDOE website includes school year, percent free and reduced lunch, total enrolment, enrolment by grade level (pre-K through 12), number and proportion English Language Learners (ELL), number and proportion of students with special educational needs (SEN) or disabilities, numbers of students with special educational needs specifically in Collaborative Team Teaching (CTT) and self-contained classrooms, number and proportion of students by ethnicity (Asian, Black, Hispanic and White), and number and proportion by gender (male and female); these were aggregated citywide and by school, district and borough. The version of this worksheet published in a given year included only schools open in that year, so worksheets published in multiple years (2010, 2011, 2012) were used to provide information for as many of the schools in the multi-year sample as possible.
- ***Network affiliation records (school level)***: For 2010-11 through 2013-14, spreadsheets were obtained listing each school's DBN, school name, management category (DOE, charter, community organization), type (general academic, special education, alternative, transfer school, home school), level (early childhood, elementary, junior high-intermediate-middle, high, secondary, K-8, K-12), grades, final grades[7], school opening date, status (open or closed), address, principal name, principal title (e.g. 'interim acting'), principal contact information (phone, fax), district, superintendent name, cluster number, cluster name, cluster leader name, cluster leader title, cluster leader phone number, network number, network name, network leader name, and network leader phone number.
- ***Network evaluation information (network level)***: For 2011-12, a document containing network evaluations is publicly available from the NYCDOE

[7] Schools in the process of phasing in or out or transitioning (e.g. from K-8 to K-12) had different values for which grades were presently served versus which grades would eventually be served.

website, although explanations to the included measures are not included. Ratings include mean Progress Report percentiles and quartiles, a qualitative indicator (1-4), Quality Review mean percentiles and quartiles, and principal satisfaction survey score quartiles and percent by response (1 item on a 4-point Likert scale, 1=”Very Satisfied” to 4=”Very Dissatisfied”). Unfortunately, comparable documents were not available for other years.

Table 1 Data sources and availability

Source	Availability/Procedure to obtain
Biographical records (student)	Formal request to NYCDOE RPG
Assessment records (student)	Formal request to NYCDOE RPG
Assessment performance (school/district/city)	Publicly available for download
Demographic snapshots (school/district/city)	Publicly available for download
LCGMS school data (school/network)	Informal request to RPG contact
Network evaluation data (network)	Publicly available, limited clarification

SPSS (IBM Corp., 2012) was used to import the various files containing these records from Excel and text formats, and to create working databases by matching on scrambled student identifiers at the student level, and school DBN (and year, as appropriate) at the school level. Network evaluation data were considered separately from the other data sources; these did not contribute to the findings presented in this document, but serve as a basis for comparison for the quantitative findings in Chapter 5.

It is important to note that despite the large amount of available data, the variability in quality and completeness of these data imposed several notable limitations. For some variables, such as assessment scores, technical reports existed to provide extended measurement information. For others, such as free and reduced lunch eligibility, it was impossible to determine the accuracy of classifications. Missing values, primarily at the student level but also for some school-level variables, provide another cause for concern, which is addressed in more depth in Section 4.5.3.

A further limitation concerns the apparent lack (at the time of the data request) of available, reliable teacher or classroom data, despite the fact that recent EER literature highlights the likelihood of significant teacher and classroom effects. This may provide a direction for further inquiry after the completion of this research, but at present it is acknowledged as a reason to interpret with caution, understanding that some of the apparent variation at the school level, for example, may in fact be attributed to teacher or class effects. While the statistical effects of teachers cannot be addressed in this study, the qualitative strand of the research explores teachers' perspectives, which may help to mitigate the omission of the teacher level in the quantitative strand.

Despite these limitations, the linked datasets provide information at different levels of the school system, allowing for the sophisticated statistical approaches required to address the relevant research questions (RQ1 and RQ3) that focus on network, district, and school-level effects on student achievement.

Table 2 gives the operational definitions of the variables used for analyses in Chapters 5 and 7.

Table 2 Operational definitions of variables

Variable	*Definition*
Literacy	8th grade literacy (English Language Arts) scale score, standardised over the population by year
Math	8th grade mathematics scale score, standardised over the population by year
Prior literacy	5th grade literacy (English Language Arts) scale score, standardised over the population by year
Prior math	5th grade mathematics scale score, standardised over the population by year
Gender	Categorical variable for student gender (0=female, 1=male)
Ethnicity	Categorical variable for student ethnicity (0=White, 1=Black, 2=Hispanic, 3=Asian, 4= Native American, 5=Multi-racial)
FSM	Indicator of free or reduced lunch eligibility (0=ineligible, 1=eligible)
Age	Student age at assessment, computed from year and month of birth, standardised over the population by year
English Proficiency	Indicator of English Language Learner (ELL) status (0=English proficient, 1=ELL)
Disability	Indicator for students with disabilities (analogous to special educational needs) (0=no disability, 1=disability)
School mean intake literacy	Aggregated literacy intake measure, calculated based on current 8th graders' 5th grade (standardised) scores by school and centred on the citywide mean
School mean intake math	Aggregated mathematics intake measure, calculated based on current 8th graders' 5th grade (standardised) scores by school and centred on

	the citywide mean
Internal/ External	Indicator of whether a network is associated with an external Partnership Support Organisation (PSO) (0=internal NYCDOE, 1=PSO)
Network elem/ middle/high	Three separate variables indicating the number of each type of school in a network
Network time	Categorical variable indicating the number of years a school has been with its current network by a particular year
Network size	Number of schools in a network (total)

4.5.2 Sample

As noted above, analyses focused on four academic years from 2009-10 to 2012-13. As the specific network structure studied was only fully implemented in 2010-11, these years provide a baseline and three additional time points.

Choosing a particular level of schooling was deemed advisable to facilitate comparisons between schools. Middle grades in particular were chosen for several reasons. First, this allowed for a choice of prior attainment measure on which most students were likely to have been assessed. All students are expected to take state assessments in grades 3 through 8, so defining prior attainment for an elementary school outcome measure would have been problematic. Second, middle school assessments are administered more uniformly and comparably than are high school examinations; the latter would lead to a variety of problems in analysis and interpretation because, in the years considered, students did not necessarily take particular exams at the same grade level within or between schools. Finally, middle school marks a crucial period in a student's academic career in New York City. Up to 8th grade, school enrolment most often depends on a student's home neighborhood. High school enrolment, however, is more complex and less geographically-defined. An application and selection process involves matching students' ranked school preferences to schools' preferred selections of students, and high-achieving students are well-placed to be selected as top choices by their preferred schools. Additionally, some high schools have specific achievement entry requirements. A student's performance in middle school, therefore, may have powerful implications for that student's subsequent educational trajectory. All of these aspects made middle grades a feasible and interesting, as well as important, level of schooling on which to focus.

Several considerations were taken into account to define inclusion and exclusion criteria for individual records and groupings of records (i.e. schools) in the sample used for

analysis. Only public schools serving (but not necessarily limited to) middle school grades 6, 7, and 8 were considered for the purposes of this study. Schools serving non-typical populations (dedicated special education schools and "transfer" schools serving only overage students) were omitted, as comparisons between these and general education schools would be unreasonable. Schools with very low numbers of records in a given year (fewer than 10) were omitted, as this indicated possible phase-out or data entry error (public schools would be unlikely to enroll such low numbers for an entire grade level).

Students who were retained in or accelerated through one or more grades during the relevant period were omitted, as values imputed for these students, based on the data for students with standard grade promotion histories, would be highly questionable. Similarly, students recorded as having taken grade-inappropriate tests (i.e. students for whom the grade levels listed in biographical and assessment records did not match) were omitted as this may indicate mis-coding (there were fewer than 50 of these for each subject).

The size of the final sample, after applying the above inclusion-exclusion criteria, is listed in Table 3 by year and level (students, schools, and networks). After the missing data treatment described below, the sample was balanced across the two achievement outcomes (i.e. equal numbers of students, schools and networks were used for analysis).

Table 3 Final sample size (students, schools, networks) by year

Year	N (Students)	N (Schools)	N (Networks)
2009-10	70925	453	Not applicable
2010-11	69216	467	57
2011-12	68023	477	57
2012-13	66619	483	56

Tables 4 provides descriptive statistics and frequencies by year for the student-level variables in the original dataset prior to missing data treatment and computed based on variables in their original scales (e.g. prior to transformations such as centring and standardisation). Most values appear fairly stable between years, but there are some apparent differences by year in the assessment descriptive values that raise some question of comparability between years to examine school level change, despite the fact that

technical reports for these assessments assert that scores are comparable across years at the same grade level. A further complication arises from the change of assessments in 2012-13, which had a different scale from previous assessments. This was one of the reasons for score standardisation (to help inform more meaningful comparisons) as well as the necessity of modelling each year's data separately. The FSM proportion is also noticeably lower in 2012-13, suggesting that although the data guides received for this dataset showed the same definition for this variable as did the guides for previous years, there may have been a change in recording procedures (e.g. a change from a free/reduced lunch indicator to only a free lunch indicator).

Table 4 Descriptive statistics by year

	2009-10				2010-11				2011-12				2012-13			
Variable	**Min**	**Max**	**Mean**	**SD**	**Min**	**Max**	**Mean**	**SD**	**Min**	**Max**	**Mean**	**SD**	**Min**	**Max**	**Mean**	**SD**
Literacy score	430	790	652.03	27.81	430	790	650.83	21.21	430	790	654.03	20.59	131	417	293.98	34.50
Math score	480	775	674.54	34.18	480	775	675.49	34.72	480	775	678.34	30.30	119	403	299.36	36.69
Prior literacy score	495	795	657.35	36.23	495	795	662.13	29.91	495	795	670.12	32.21	495	795	671.15	31.48
Prior math score	495	780	672.68	38.71	495	780	679.25	37.53	495	780	686.21	32.93	495	780	685.43	32.89
Age	11.58	18.83	13.99	0.61	11.50	18.58	13.97	0.59	10.58	18.25	13.94	0.57	11.58	18.33	13.91	0.54
	Freq	**%**			**Freq**	**%**			**Freq**	**%**			**Freq**	**%**		
Gender (Male)	35990	50.7			35278	51.0			34367	50.5			33959	51.0		
Ethnicity (White)	9842	13.9			9479	13.7			9667	14.2			9684	14.5		
Ethnicity (Black)	21595	30.4			20280	29.3			19614	28.8			18403	27.6		
Ethnicity (Hispanic)	27940	39.4			27786	40.1			27275	40.1			26928	40.4		
Ethnicity (Asian)	11193	15.8			11211	16.2			10973	16.1			11052	16.6		
Ethnicity (Multi)	86	0.1			153	0.2			362	0.5			160	0.2		
Ethnicity (Native American)	218	0.3			306	0.4			130	0.2			385	0.6		
FSM (eligible)	60872	85.8			59513	86.0			56798	83.5			49538	74.4		
Eng. Proficiency (ELL)	8296	11.7			8888	12.8			8667	12.7			8674	13.0		
Disability	10980	15.5			11046	16			10956	16.1			11263	16.9		

4.5.3 Missing data

The importance of handling missing data appropriately is widely acknowledged in quantitative research (Enders, 2010). While initial exploratory analyses were the result of list-wise deletion, which simply excludes any incomplete records and has been widely used in multilevel modeling (Goldstein, 2011), analyses presented in this thesis are based on the use of appropriate strategies to handle missing data. This section examines missingness in the datasets used in this study, and draws on the relevant literature to justify the appropriateness of the strategy used for missing data treatment (multilevel multiple imputation).

Van Buren (2015) outlines a list of considerations to be used in classifying missing data, which is a necessary step before determining appropriate procedures to apply. These considerations are as follows:

- Role: Which variables are incomplete (response and/or predictor(s))?
- Pattern: In what, if any, patterns are data missing?
- Mechanism: How, if at all, is the probability of a value being missing on a particular variable related to the data?
- Design: What is the design of the study (random, clustered, longitudinal)?

Each of these has implications for the treatment of missing data. These considerations are discussed in the sections below.

4.5.3.1 Missing data roles and design

The design of the study is clustered, due to the inherently nested structure of the data. This is an inherent feature in the study of potential effects of schools and higher groupings (e.g. districts and networks). In the datasets used for this project, there were missing values in the dependent variables (literacy and mathematics assessment scores) as well as in many of the predictor variables, and there were missing values at the student and school levels (some schools were missing demographic or attainment data in the spreadsheets provided on the NYCDOE website). Rather than automatically imputing missing school data, further efforts were made to find school-level missing values as appropriate by looking up the original Progress Reports and Quality Reviews containing the relevant information on school websites; removing these records, or imputing values if there was a substantive

reason for a missing value not immediately obvious within the dataset, would have risked biasing the results unnecessarily (Ibid). Several schools were removed rather than having values imputed when they were found to be "phasing out" (closing or changing grade configuration) and not subject to reporting in a given year; this was judged to be acceptable as such schools were arguably not comparable to schools serving the full range of middle grades. After researching and filling in values where this could be done reliably and reasonably and removing phase-out schools, there were relatively few remaining missing values at the school level.

4.5.3.2 Missing data patterns

Patterns of missing data can be characterised as univariate (with only one incomplete variable), monotone (where subsequent variables are missing after one variable is missing in a particular order, as in a longitudinal survey design with dropout), general (seemingly haphazard), latent variable (where values of a variable are missing for an entire sample), planned missing (common in survey methodology), or unit-nonresponse (as in a survey with a question participants omit) (Enders, 2010; Little & Rubin, 2002; van Buren, 2011). Several of these are unlikely to be relevant in the present context, as they are specific to survey studies, and as more than one variable is clearly missing, the pattern is clearly multivariate and not univariate. The following discussion considers whether the pattern is monotone or general.

As acknowledged in the explanation of sample inclusion and exclusion decisions, some of the data missing from the relevant datasets was missing for substantive reasons. For example, some schools were not opened until 2012-13, or were closed after the end of 2009-10. Thus, these schools would not have had any students in certain years, and imputing values for those (nonexistent) students and schools would be inappropriate. Similarly, some networks were eliminated or initiated from year to year. Therefore, missing values were analyzed separately for each year's data (to avoid this issue of entire missing schools or networks).

In total, approximately 80% of the cases in the dataset for each year were complete. For 2009-10 and 2010-11 there were missing values for 7 of 16 variables included in the missing data analysis, while for 2011-12 and 2012-13 there were missing values for 10 of 16 and 9 of 16, respectively. The proportion of total missing data for each year in terms of

cell entries (number of cases multiplied by number of variables) was low: 2.4% for 2009-10, 2.5% for 2010-11, 3.0% for 2011-12, and 2.8% for 2012-13.

A comparison of missing data patterns across school- and student-level variables was undertaken to determine whether variables appear more likely to be missing jointly in particular combinations. Appendix 1 provides a comprehensive account of missing data frequencies and patterns as well as distribution by variable and case. From these, there does not appear to be a monotone pattern of missing values in any given year.

4.5.3.3 Missing data mechanisms

Missing data treatment choices are largely dependent upon the mechanisms behind the missing values. Missing data mechanisms are classified as belonging to three possible categories (Little & Rubin, 2002):

1) Missing completely at random (MCAR), in which the probability of a value being missing is not dependent on values of the unobserved nor observed data;
2) Missing at random (MAR), in which the probability of a value being missing may depend on observed data but not on the missing value itself; and
3) Missing Not at Random (MNAR), in which the probability of a value being missing depends on the unobserved data (i.e. the missing value itself).

It is only possible to prove that data is not MCAR; without knowing the missing values, it is impossible to determine conclusively whether data is MAR as opposed to MNAR (Ibid). The test for MCAR is a multivariate chi-square test, taking the overall mean obtained by expectation-maximization (J L Schafer, 1997) and comparing this to the observed mean for each missing pattern (Little, 1988). Results from running this test on school and student variables for each year are given in Appendix 1; the test statistics are clearly significant, signifying that the data are not MCAR, but are taken to be MAR (Little & Rubin, 2002). Implications for data treatment are discussed in the following section.

4.5.3.4 Missing data treatment

List-wise deletion is only entirely appropriate for data that is MCAR, so that the list-wise deleted dataset is a representative sample of the complete original sample (Enders, 2010; J L Schafer, 1997; van Buren, 2011). Where data is not MCAR, list-wise deletion runs the risk of biased estimates. For multilevel models, recent developments in multilevel multiple imputation provide an advanced approach to missing data treatment. In most instances it is

essential to account for the multilevel structure of the data in the imputation model, as "flat-file" imputation (using dummy codes for higher-level clusters and treating the data as a single level) can lead to biased estimates based on simulation studies (e.g. van Buren, 2011). This is particularly important to consider when cluster effects, or in the present study school and/or network effects, are of substantive interest (Drechsler, 2015).

Multilevel multiple imputation has been suggested by numerous authors as an appropriate method to treat missing data that is MAR (van Buren, 2015). Given the patterns and mechanisms of the missing data in this study, together with the substantive focus of the proposed analysis and research questions, this approach was an appropriate choice. Multiple imputation computes plausible values to replace missing values by using additional variables with no missing values (auxiliary variables) from a given dataset as predictors. This process is repeated to create multiple complete datasets. Following the imputation process, analysis is conducted on each imputed dataset and then results are combined or "pooled" according to Rubin's Rules (Rubin, 1987). Benefits of multiple imputation include the fact that is uses observed information to handle non-ignorable missingness, and the creation and analysis of multiple datasets avoids understating uncertainty, which is a concern for alternative single-imputation approaches to missing data treatment (J L Schafer & Graham, 2002). MPlus (L. K. Muthen & Muthen, 2015) was used; this is one of several software programs that have recently been developed or extended to handle multilevel structures. Such programs are limited to at most two levels for imputation models, so in this instance two-level models were used with students nested in schools.

Recommendations regarding the optimal number of imputations vary widely. One approach taken by many authors involves calculations of estimation efficiency, wherein if λ is the rate of missing information and m is the number of imputations, then efficiency is calculated as $(1+ \lambda/m)^{-1}$. On this basis, even the most conservative calculation replacing λ with the highest proportion of missing values on any variable (approximately 0.2), it would seem that as few as five imputations would lead to an estimation efficiency of 96%. Schafer and Olsen (1998) indicated that the gains in efficiency associated with more than 3 to 5 iterations are likely to be minimal. However, Graham et al. (2007) demonstrated through a simulation study approach that in practice many more imputations are needed to avoid loss of power, and suggested that in fact standard errors and p-values can also vary considerably based on the number of imputations. Based on the recommendations of these

authors, using a conservative estimate of λ=0.2, 20 imputations were used in the present study for each separate year for which analysis was conducted. Running 20 imputations for each year and then analyzing these for each model did take considerably longer than undertaking the same imputation and analysis process for 5 or 10 imputations. However, the cost in terms of time was mitigated somewhat by running multiple models (or iterations of a particular model) simultaneously in MLwiN, which required approximately twice (rather than four times) the amount of time required for 5 imputations, and a larger number of iterations did not substantially complicate the calculations involved in pooling estimates. The potential loss of power associated with a smaller number of iterations, along with the concern that estimation efficiency alone may not provide sufficient justification for a particular number of iterations as noted by Graham et al. (Ibid.), supported the decision to use 20 imputations and justified the additional time associated with this choice.

4.5.4 Quantitative analysis: Multilevel modeling

In the context of the present study, students are grouped within schools, which in turn belong simultaneously to geographically defined districts as well as to networks of schools opted into by each school's principal on an annual basis. Given these relationships between pre-existing 'levels' or clusters of the school system, the assumption of independent and identically distributed variables necessary for traditional regression approaches is unlikely to be met, because students in the same school may be more similar by some measures than students from different schools. The same may be true for schools in the same district or network as compared to schools in different districts or networks. Thus, because of the underlying structure of the data, and the particular interest of the study in examining potential school and network effects, multilevel analysis was selected as the most appropriate way to model relationships and variation (Goldstein, 2011; Hox, 2010; Raudenbush & Bryk, 2002; Snijders & Bosker, 2012).

Of particular interest for this study are intra-class correlations (ICC), which express the overall proportion of total variation in student outcomes attributable to the "effect" at a particular level. For example, the ICC at the school level can be interpreted as the correlation between (the scores of) two students within the same school, so that the ICC corresponding to each level in the model measures the proportion of total variance accounted for by clustering at that level. Residual estimates and their confidence intervals are also substantively important, particularly for individual schools or networks, partly

because of their role in sampling strategy for the qualitative strand (see Section 4.5.5). In multiple regression, a residual is merely the distance from the regression line to an observed value. In multilevel modelling, residual calculations are slightly more complex, as cluster size and variation must be factored in higher levels of the model; formulas guiding these calculations for the purpose of the present study are presented in Section 4.5.5.

Before any multilevel analysis was undertaken, the first stage of quantitative inquiry was a thorough exploratory descriptive analysis using SPSS (IBM Corp., 2012). This included studying the distribution of student performance in literacy and math, overall as well as according to different variable values, to identify any apparent relationships between the outcome variables and various student, school, district and network characteristics. Selected findings from this preliminary analysis can be found in Section 5.4.

Model fitting follows a general approach suggested by Hox (2010) and by Snijders and Bosker (2012), beginning with an empty, single level model and adding levels, fixed effects, and random terms individually and sequentially so that the significance and improvement to model fit can be checked at each step of the process. All of the models are fitted using MLwiN 2.31 (Rasbash et al 2014), as this software is uniquely able to handle – and produce useful graphics for – complex multilevel models.

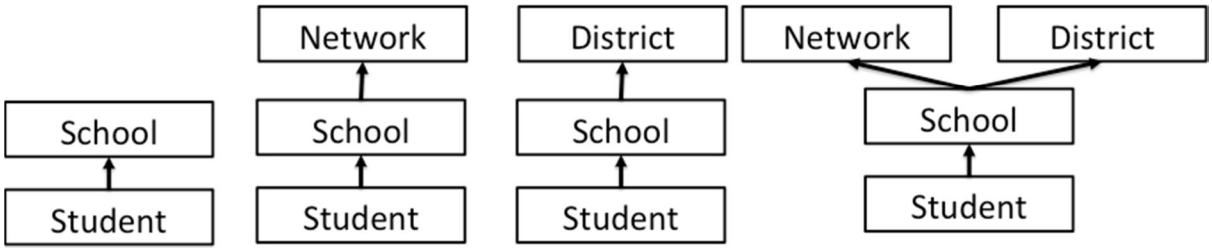
Because, as noted in the technical reports for New York State ELA and Math assessments (CTB/McGraw Hill, 2010a, 2010b, 2011a, 2011b, Pearson, 2012a, 2012b), assessments are not vertically equated across grade levels, direct modelling of individual student growth was unfeasible (Singer & Willett, 2003). This precluded the use of what would otherwise have been preferred methods of analysing change over time, most notably multilevel growth modelling. As the data would not allow for more sophisticated longitudinal analysis approaches, model comparisons across years and examinations of patterns in school performance categories over time were used to explore change in achievement over time.

In estimating between-school or between-network variance, the question of what controls to include is an important one. The dynamic model (Creemers & Kyriakides, 2008) cannot be fully tested or applied in this study because variables at some of the relevant levels (teacher, classroom) were not available at the time of request. However, the dynamic model highlights the consideration of school effects and the inclusion of student

background variables and school contextual and compositional variables. It is worth noting that composition and context were not calculated as aggregates of student-level values, instead using the publicly-available school-level data to avoid exacerbating measurement or sampling error through aggregation. The one exception to this was school-level prior attainment, which was calculated (prior to list-wise deletion) as aggregates of student prior attainment because no such variable was available at the school level from the NYCDOE.

There is some controversy in the literature about whether school effects should be interpreted based on models including or excluding student background characteristics. A school that appears to have a large effect after controlling for ethnicity, for example, might not be performing well overall, and treating that school as "high-performing" for the purposes of further qualitative study would be highly problematic. On the other hand, it is important to investigate variance explained at each level by such characteristics and the resulting residuals, because this gives insight into how much of an observed school effect is actually driven by differences between schools in their intake compositions. The same questions can be extended to network or district effects. For these reasons, the following chapter presents and compares results from several models, including empty (unadjusted) models, models adjusting only for prior attainment, models adjusting for a range of student-level predictors, and finally, models adjusting for school-level variables including composition, context and network affiliation characteristics. Attention is paid to comparing the ICCs, and individual residuals of schools and their associated confidence intervals, between these various models for each subject.

Figure 3 Classification diagrams: 2-level, 3-level, and cross-classified models



Analysis proceeds through the following steps (separately for each subject), which collectively address RQ1 and associated sub-questions:

1) Comparison of differently specified models to explore the variance structure (see Figure 3 for classification diagrams of these models). Because models are not all strictly nested within each other, the Deviance Information Criterion is used to

compare model fit (Goldstein, 2011). Subsequent analysis then proceeds with the model(s) selected for best fit. Models compared include

a) students-within-schools (2-level),
b) students-within-schools-within-networks (3-level),
c) students-within-schools-within-districts (3-level), and
d) students-within-schools cross-classified within networks and districts (4-level or 4-classification)[8].

2) Inclusion of student prior attainment
3) Inclusion of student background variables
4) Inclusion of school-level context and composition, network-related variables, and consideration of cross-level interactions
5) Repeating steps 2-4, modelling of school-level change in attainment

The complexity of some of the models fitted in this analysis made the choice of an appropriate estimation method crucial. Although there are documented procedures to estimate cross-classified models in MLwiN using maximum likelihood estimation methods, these may lead to biased estimates. Markov Chain Monte Carlo (MCMC) estimation in MLwiN (Browne, 2015) was chosen instead as the most appropriate approach for the present study. Adjustments were made as needed to the default MCMC settings in the software, as typically the models required longer chains and larger burn-in values[9]. Hierarchical centring, which is given as an option for MCMC in MLwiN, was employed as a way to improve mixing at higher levels of the models, but because of the way in which this procedure is automated in the software, this does not change the interpretation of results.

Correlations between higher-level residuals from models in different years and subjects (containing the same schools) were calculated as measures of stability over time and consistency across subjects.

For the analyses presented in this document, adjustments have not been made for measurement error due to time constraints and unwieldy technical processing requirements of combining measurement error adjustment with MCMC estimation and multiply imputed data. However, previous studies suggest that ignoring measurement error may lead to

[8] Strictly speaking, in cross-classified models the "levels" are referred to as "classifications" in order to acknowledge that they are not strictly hierarchically nested.
[9] Burn-in refers to a specified number of initial iterations that are discarded at the beginning of an MCMC chain.

biased estimates of various predictors (Battauz, Bellio, & Gori, 2011; Televantou et al., 2015; Woodhouse et al., 1996); accounting for measurement error would be a possible direction for extending the work in this thesis in later publications.

4.5.5 Informing the selection of schools for qualitative study

Because of the time constraints of the doctoral course, sampling for the qualitative strand of the study was based on contextual value added models (presented in Chapter 5), specifically those estimated for the normalized achievement results in 2012 in literacy and mathematics. The limitations of sampling based on a cross-sectional measure of school effectiveness are undeniable, but lacking the appropriate data for common approaches to longitudinal modeling, the approach taken was seen as the best option under the circumstances to support purposive sampling. Previous empirical studies offer precedent for the use of multilevel model residuals within MM designs, both to characterize cases (e.g. Day et al., 2008) and to inform sampling (e.g. Sammons et al., 2005). The sampling strategy presented here adapts a method used in the Effective Provision of Pre-School Education (EPPE) project (Ibid.).

In order to categorize schools as high-, typical-, or low-performing, the shrunken residual ($\widehat{u_j}$) for each school was calculated according to the formula[10]

$$\widehat{u_j} = \frac{{\sigma_u}^2}{{\sigma_u}^2 + \frac{{\sigma_e}^2}{n_j}} r_j$$

where r_j is the raw residual for school *j*. Variances for these effects were also calculated as follows:

$$var\left(\widehat{u_j} - u_j\right) = \frac{{\sigma_u}^2 {\sigma_e}^2}{n_j {\sigma_u}^2 + {\sigma_e}^2}$$

to enable the calculation of a 95% confidence interval for each school,

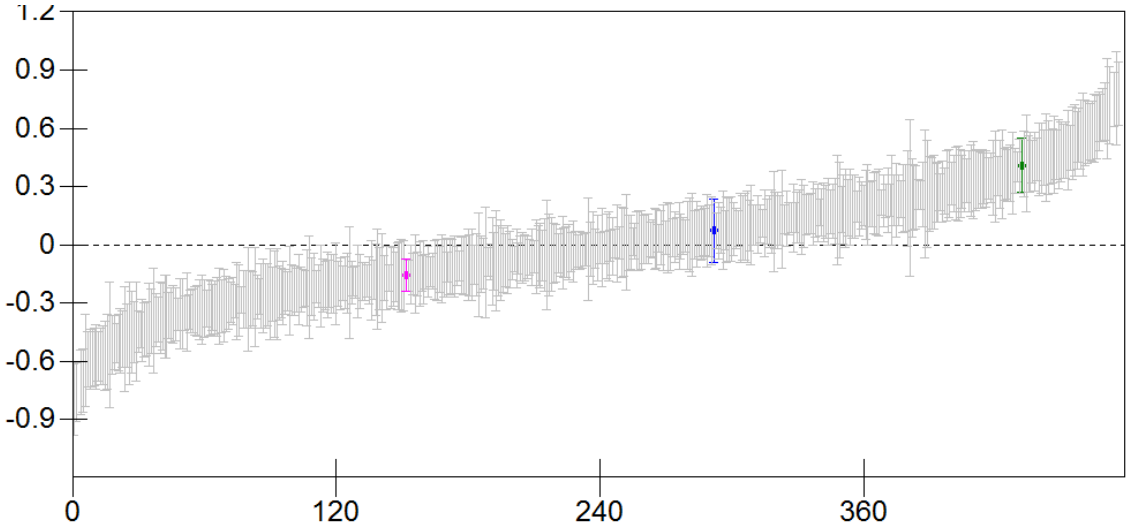
$$\widehat{u_j} \pm 1.96\sqrt{var\left(\widehat{u_j} - u_j\right)}$$

The residual estimates, along with their confidence intervals, allow for comparisons between individual schools as well as between groups of schools by network. For the

[10] Note that the software used, MLwiN, performs this calculation automatically; formulae are provided for the sake of replicability regardless of software choice.

purpose of sampling in this study, schools were classified as high-performing if the estimates of their effects were significantly above zero (including 95% confidence intervals), and low-performing if significantly below (including 95% credible intervals). An illustration of ranked residuals with their confidence intervals is given in Figure 4 for the 2012 2-level literacy model, with a high-performing school highlighted in green, a low-performing school highlighted in magenta, and a typically-performing school highlighted in blue.

Figure 4 Illustration of high, typically and low performance categories based on residuals and confidence intervals



4.6 QUALITATIVE DATA, SAMPLE AND METHODS

Overall, qualitative inquiry in the present study can be described as taking a case study approach, designed to address RQ2 and its associated sub-questions. As in any case study, it is important to define what is meant by a "case"; over the course of this research project, that definition has changed. Initially, the intention was to pursue an embedded multiple-case study in which networks were the main units of analysis and multiple schools nested within each network served as embedded subunits of analysis (Yin, 2014). In practice, access challenges and complications (see Sections 4.6.2 and 4.10) made this impractical. Thus, the study has been reframed as a single-case study, in which the case is redefined as the entire network structure in the NYCDOE, and data collected from schools, networks and participants within that structure help to inform understandings about the network structure as a whole. Schools serve as subunits of analysis, and also as cases within the main case study, so that comparisons across schools can shed light on the network structure as a whole (Ibid). The boundaries, too, have shifted from processes, interactions,

individuals and organizations within several specific networks, to the same list of subjects, objects and interactions across the entire network structure as illustrated by embedded school cases. Simons (2009) suggests that this redefinition is often a natural part of the case study process, stating that boundaries may "shift in the course of conducting the study and once you begin the analysis" (p.29).

4.6.1 Why case study?

Stake (1995) claims that "Case study is the study of the particularity and complexity of a single case, coming to understand its activity within important circumstances" (p.xi). Bassey (1999) offers a more extended definition of case study as an empirical enquiry which

- is temporally and spatially bounded
- focuses on "interesting aspects" of an educational institution or system
- is conducted in natural context and with an ethic of respect for participants
- is intended to inform practitioners, policymakers or theoreticians
- involves collecting enough data to explore significant features, inform plausible interpretations, construct and convey a worthwhile and convincing narrative, relate findings to existing literature, and enable replication and argument by other researchers.

This understanding of case study, its purpose and potential, forms the rationale for the methodological choices in the qualitative strand of the present study. Taking into consideration the research questions and practical characteristics of the setting, a case study approach was deemed most appropriate because it allows for the in-depth study of a bounded phenomenon of interest in its real-world context, and fits well with the use of mixed methods and mixed data (Yin, 2014). Experimental methods would have been inappropriate, as networks are intended to vary in approach and scope and no intervention was being delivered; this research studies naturally occurring variation (among schools and networks). Survey methodology would have been a plausible approach, but would have limited potential findings to the items included in the survey, whereas case study methodology facilitated a more in-depth and exploratory approach as well as the potential for unanticipated themes and topics to emerge.

4.6.2 About the sample and sampling process

The overall approach to sampling can be described as sequential MM sampling, in which samples for the quantitative and qualitative strand were defined at different stages of the project. Multiple purposive sampling was undertaken within the qualitative strand, as samples for various levels of the system (school, teachers, networks) were approached differently due to the complexities of the setting (Teddlie & Yu, 2007).

Section 4.5.5 described the process by which quantitative methods were intended to inform the selection of schools for qualitative case study. While this procedure was used to determine which schools to recruit, in practice the final sample was less systematic than was intended in the abstract. Ultimately, the final sample consisted of ten schools in a variety of networks. The numbers of participants in each stakeholder group are given in Table 5.

Access is much discussed in the case study literature, and as is typical for this type of design, the present study involved getting access to schools through the principal as a gatekeeper (Simons, 2009; Yin, 2014). Because there were multiple steps involved in gaining access to schools, and because participation was voluntary at the individual level (for more detail, see Section 4.11), not all schools had corresponding participants in every group. The principal was approached first in each school, because contact information for these individuals (email and phone) is readily and publicly available, and because the NYCDOE IRB requires principals' formal permission to be obtained (via a standard form, included in Appendix 2) before any research takes place in a particular school. A principal's participation and consent, however, did not guarantee the participation of corresponding network staff or teachers. The original design included plans for sampling teachers in multiple departments in each school, but in practice this was not always possible based on the response to recruiting efforts (in one school, the only teachers who responded were in the same department as one another).

Table 5 Numbers of participants by stakeholder group

Type of participant	Count	Number of schools including participant type
Principals	10	10
Teachers	17	5
Network staff	4	3
Total	31	10 (Total schools)

All data collected was included in the analysis despite this variability in participation from school to school; imperfect or incomplete data was deemed more valuable than complete data in a much smaller sample. While the actual sample was more opportunistic than intended based on the original research design, findings are still meaningful, provided that potential bias arising from sampling issues is attended to during analysis and made explicit in reports of results to avoid a loss of rigour or credibility (Yin, 2014). A breakdown of the sample of schools by performance category in each subject is presented in Table 6, and an illustration of schools' effects relative to all schools included in the quantitative analyses is given in Figures 5 and 6 (with sample schools highlighted).

Table 6 Cross-tabulation of schools in the sample by performance category and subject

		Mathematics 2011-12			
		Low	Typical	High	Total
Literacy 2011-12	Low	1	1	0	2
	Typical	5	1	1	7
	High	0	1	0	1
	Total	6	3	1	10

Figure 5 Residuals/confidence intervals, effects of sample schools: Literacy 2011-12

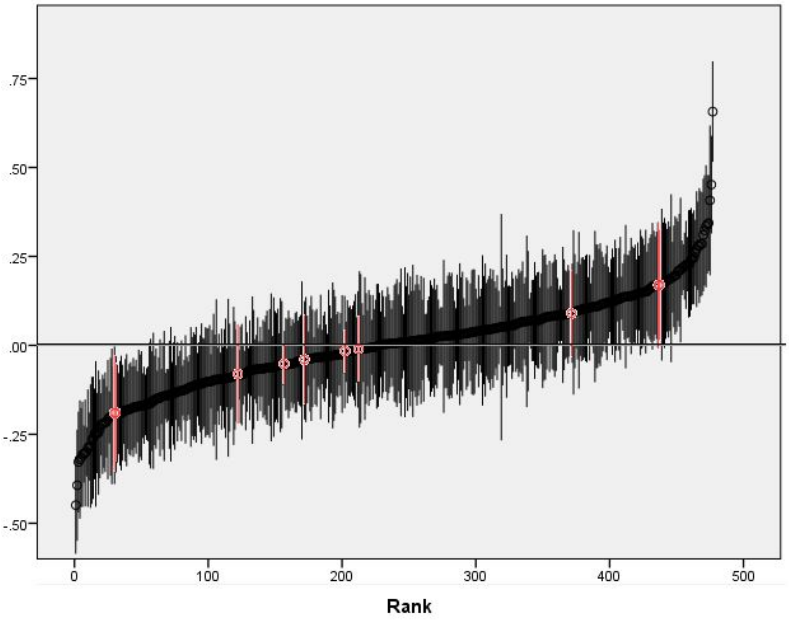
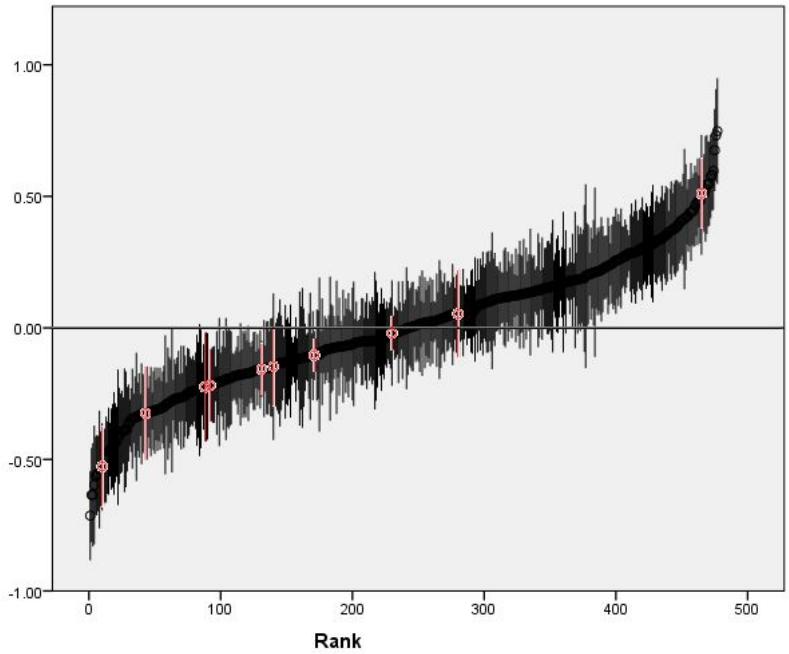


Figure 6 Residuals/confidence intervals, effects of sample schools: Mathematics 2011-12



4.6.3 Qualitative data collection/sources

Yin (2014) proposes that case study approaches are not limited to any specific set of methods, but that multiple sources of evidence are crucial. The original research design for this study included two of the most common methods and sources used for case studies: interviews and observation (Merriam, 1988; Simons, 2009; Stake, 1995). An adjustment to

this design was necessary, because it became clear early in the fieldwork process that many participants were not comfortable with the idea of my observing meetings. That, added to the difficulty accessing networks directly (and therefore being able to access network activities and meetings), made it prudent to remove observations from the design in the interest of minimizing harm and maintaining consistent methods across schools.

4.6.3.1 Interviews

For each school case included in the sample for the qualitative strand of the study, interviews were conducted with the school principal and, depending upon participants' willingness to participate, various teachers and staff from the network affiliated with the school.

Rubin and Rubin (2012) suggest that interview methods are appropriate when the processes being studied are invisible or inaccessible to the researcher. This suits the current purpose, as I was not privy to the interactions and processes operating within networks first-hand. Perhaps a more poetic justification for the use of interviews is that they are "the main road to multiple realities" (Stake, 1995, p.64) or, in other words, the perspectives of multiple individuals and multiple groups of stakeholders identified in the research questions for this study.

Interviews were semi-structured, guided by an interview schedule that was framed as similarly as possible for the different stakeholder groups in order to facilitate comparisons. The choice of semi-structured interview was an intentional one. Unlike structured survey methods, semi-structured interviews allow for in-depth responses and explanations, and can be undertaken flexibly and adaptively so that the wording and questions may change to fit the needs of the specific situation and conversation (Patton, 1980; H. J. Rubin & Rubin, 2012). By comparison to unstructured interview, however, this method allowed the interview to remain focused on the topics of interest. There were also practical considerations; many of the interviewees were busy and reluctant to participate unless I could promise specific time constraints. Semi-structured interview methods allowed me to address this issue appropriately (Simons, 2009).

The ordering of questions was largely determined by direct relevance to the research questions, and by feedback from participants in pilot interviews (see Section 4.6.3). This was done so that in case of an interruption or necessity to end the interview before completion, the questions considered most fundamental to the main lines of enquiry would

still have the best chance to be addressed. The exception to this rule was that all interviews began with introductions (to the project and to me) and "easy questions" about participants' backgrounds (H. J. Rubin & Rubin, 2012).

Topics covered in interviews included:

- Information about participants (job role, career history, time in current organization)
- Information about organizations (school context, network context)
- Change in the school or organization (since recent network affiliation or since before the network structure)
- Improvement/development priorities and planning/evaluation processes
- Organizational culture
- Interactions between schools and network organization
- Interactions/collaboration between schools or school staff
- Leadership and decision-making processes
- Impact of network involvement and/or collaboration between schools
- Reactions to external pressures/initiatives/policies

There is some conflict in the methodological literature about whether to audio-record interviews (e.g. Rubin & Rubin, 2012; Simons, 2009; Stake, 1995). All advice seems to converge on one point, which is that interviews should be recorded meticulously and accurately, and records of interviews should ideally include notes about tone, relationship between interviewer and interviewee, and the researcher's experience in the interview. In the interest of fidelity to participants' own words, interviews were audio-recorded except in several instances in which participants preferred to opt out of this (in which case diligent notes were taken). I also endeavoured to take notes during the interview to document aspects not captured by the audio-recording and to note key ideas and relational dynamics where applicable. Transcription took place as soon as possible after each interview, usually on the same day or the following day.

4.6.3.2 Piloting and timing of data collection activities

Interviews were piloted in one school before beginning data collection for the main study. The results of these interviews have not been included in the thesis, as the main purpose of the pilot was to hone the interview guide and my approach to interviewing. Pilot interviews were conducted with 7 teachers and 2 administrators (a principal and an

assistant principal), as well as one network leader. The main consequences of the pilot were changes to question phrasing and a re-ordering of subtopics based on participant feedback (some questions, especially those pertaining explicitly to networks, were perceived as easier to answer after having thought about others). The rationale for excluding pilot data from the final report is that I had previously worked in the pilot school before beginning my doctoral studies, and in fact had worked alongside most of the pilot interviewees. As such, while my pre-existing relationship with the school and its staff helped to establish access and critical feedback on my interview guide and approach, there was a chance of multiple forms of bias, both in the content of interviews and in the sample of individuals who consented to participate (i.e. people with whom I had a good working relationship were probably more likely to volunteer to contribute).

Data collection for the main study took place during 2013-14. The fairly long span of time during which interviews were conducted reflects practical considerations related to recruitment; some principals and network staff who agreed to participate answered recruitment calls or emails up to several months after my initial attempts to contact them, and the same was true when arranging follow-up visits to recruit teachers. "Rounds" refer to two main waves of recruitment and data collection, the second undertaken because the first did not lead to a sufficient sample size. These are recorded below to reflect that participants were interviewed at different times, which represents a challenge and limitation of the study (discussed further in Section 4.12). The timing of interviews is explicitly taken into account in qualitative analysis to address possible temporal patterns of responses, and to disentangle these from other themes and patterns.

4.6.4 Qualitative analysis

Yin (2014) posits that there are four main analytical strategies for case study research: the first relies on theoretical propositions and considers how the data relates to these, the second involves working from the ground up to develop theory, the third involves description of the case, and the fourth involves examining possible rival explanations for observed results. In the present study, a combination of these was undertaken. The goals of analysis were to:

1) Describe the main case in terms of the interactions and processes operating within it,

2) Develop situated interpretations (which may be viewed as highly context-driven theory) based on the data
3) Examine these processes and interactions in terms of existing theory
4) Consider alternative reasons for what is observed (i.e. school performance, improvement, and network interactions, relationships, and processes) in context.

The analytical approach was developed and improved over the course of the project. During the planning and design stage, some a priori codes were identified from the typology of networks by Muijs et al. (2011). These helped to frame the interview guide and guiding principles for analysis, and as such were considered important to take into account in the analysis of data (as represented in the analytical framework below). That is, these themes were pre-identified partly to aid in descriptive analysis, and partly to link results to prior literature on school networks (for contrast and comparison). Attention was given to the distinction between themes arising directly from the researchers' choice of questions and topics and themes emerging spontaneously from data.

The basic approach to thematic analysis draws from principles of grounded theory (Strauss & Corbin, 1998); however, there is a clear acknowledgment of pre-existing knowledge built upon readings of previous research literature as well as personal experience (the latter is elaborated upon in Section 4.7 below), as well as new knowledge emerging from the data. That is, I do not make the claim that this is "pure" grounded theory. The researcher, prior literature, and the contents of interviews themselves all contribute to shaping the analysis.

Interviews were analysed first as one large dataset, and later grouped separately (e.g. by stakeholder group or performance category) to examine convergent or divergent patterns. Analysis proceeded through the following stages, which can be thought of as progressive focusing (Parlett & Hamilton, 1972):

1) Familiarising myself with the data through reading and re-reading transcripts
2) Starting to code using broad categories, followed by progressively finer-grained codes within those categories (Coffey & Atkinson, 1996)
3) Regrouping sub-codes to examine different possible categories and groupings of themes and ideas

Part of this process involved restructuring the data by extracting content with a particular code into a new file and summarizing the contents of this new combination of data (a

procedure advised by Rubin & Rubin, 2012). The analytical framework was adjusted, developed and supplemented through this iterative process. NVivo (QSR International, 2012) was used to assist with data management, coding, and analysis using technical procedures outlined by (Bazeley & Jackson, 2013), but preliminary coding was done by hand.

4.6.4.1 Analytical framework

Descriptive codes given were adapted from a pre-existing typology of networks (Muijs, Ainscow, et al., 2011), covering aspects such as network density, time frame for affiliation, formal/informal connections, links to external organisations, and geographical location (school) and spread (network). For thematic codes, the top-level (broadest) categories were pre-determined by the topics covered in interviews in order to disentangle these from emergent themes arising from participants' responses.

4.7 POSITIONING

Toward the beginning of this chapter, this research was described as being founded upon a pragmatist stance. However, different authors have defined pragmatism and associated principles, values and beliefs in rather different ways. Maxcy (2003) distinguishes between epistemological pragmatism and methodological pragmatism, for example, and asserts that the former is a theory of knowledge while the latter is more directly concerned with the reasoning behind the selection of methods. It is not clear enough, then, to merely attach the label of pragmatism to myself and to my research. This may frame the research approach efficiently, but the reader is left without an understanding of what assumptions I may have brought to this project about the nature of knowledge and truth. Without discussing these assumptions, I risk losing credibility. It is equally important to reflect upon and make transparent my political, personal and professional background and potential biases relevant to the study's specific focus and context. To this end, in Section 4.7.2, I explore explicitly these different aspects of positioning and belonging (May, 2011).

4.7.1 Theoretical perspective and epistemological stance

Based purely on the research design, this project might be classified as a post-positivist endeavour because of the priority given to the quantitative strand, implying the existence of some underlying objective reality that can be measured and explained, if imperfectly (Onwuegbuzie et al., 2009). The use of qualitative methods is not necessarily discrepant

from this epistemological stance. In fact, the incorporation of multiple perspectives and questions about processes and interactions may be seen as a way of better approximating reality by integrating various perspectives and approaches as a sort of triangulation. The emphasis placed on participants' perceptions and experiences and the ways in which they are analysed, however, implies acceptance of the existence of a subjective, constructed reality, and the desire to explore stakeholders' perspectives and co-construct understandings of the ways in which networks were experienced. To further complicate matters, the use of Bayesian statistical approaches within the context of multilevel modelling also blurs some epistemological lines; Bayesian statistics have been commonly characterized as subjective as compared to classical statistical approaches, though this subjectivity assertion has been forcefully contested (Gelman & Shalizi, 2013).

In the face of these nuances, I have found it challenging to characterise myself as belonging to one "-ism" or other. Pring's (2000) criticism of 'false dualism' between paradigmatic worldviews, and acceptance of the idea that there can be both objective and subjective realities that can be studied in different ways, resonates with my worldview (and can also be logically aligned with pragmatism). As a natural consequence of this perspective, the analysis and findings reported in this thesis reflect some arguably objective realities (i.e. measured effects) as well as some subjective realities (i.e. participants' experiences and perspectives), relevant to different aspects of the same phenomenon.

4.7.2 Personal/professional background

Before becoming a researcher, I was a secondary mathematics teacher for approximately seven years, five of them in two NYCDOE public high schools (which were both in a particular network). This personal history has had an undeniable role in shaping my political views and priorities, my personal beliefs and values, and more pertinent to this project, my perspective as a researcher. In many cases, my background has given me the contextual knowledge necessary to make meaningful interpretations that I might not otherwise be likely to access. However, this also means that I cannot claim objectivity; I am bound to have opinions about a system in which I worked for years and about my former profession. For example, the inclusion of teachers' voices in the current study could be interpreted as a reflection of my bias, essentially based on a belief that teachers' views and experiences are important considerations in discussing higher-level policies.

Particularly in the context of qualitative interpretations, my approach to dealing with subjectivity was to adopt an explicit reflexivity, and to be as transparent as possible about my analytical process (see Mruck & Breuer, 2003).

In addition to affecting my own perspective, my teaching background had an effect on participants' responses to and relationships with me. In some cases, this was explicit (e.g. "You were a teacher, so you know…"), while in others it may have been subtler. In most cases, my having had experience working in New York City schools – and disadvantaged urban US educational settings more generally – helped to build trust. There were drawbacks, however; occasionally participants would fail to explain references because of an assumed shared language and knowledge base, so that I needed to ask for clarification regarding policy language and jargon that had emerged since my resignation. In a few cases, trust was actually more of an issue because of my background. This was particularly so for some of the teachers I interviewed who wanted to know, for example, why I had left the profession, and why I am pursuing my doctorate at such a faraway institution. I chose to answer these questions, if tacitly, subscribing to the view that "finding out about people through interviewing is best achieved when the relationship is non-hierarchical and when the interviewer is prepared to invest his or her own personal identity in the relationship" (Oakley, 1981, p.41). In the context of this research, I am an insider-outsider, sometimes more one or the other depending on the specific moment in the conversation or the research process. This could be considered part of the messiness of the qualitative research process. On the other hand, I maintain that being able to draw on both insider and outsider perspectives while interacting with data and participants is a distinct asset, allowing me to incorporate a broader range of interpretations.

In some instances, I had my own strong opinions about topics discussed in the interviews, not necessarily directly related to networks but to individual and school contexts. It was not appropriate to express those overtly while questioning participants, as the priority was to access their views and perspectives. I did, however, keep notes in a journal for myself after each interview to allow me to consider, in analyzing the qualitative data, how my perspective might be playing a role in the analytical process. Whereas in a more phenomenological study this might have figured more explicitly into the findings (e.g. Mosselson, 2010), for the purposes of this study the goal was to examine my biases in order to avoid conflating them with or imposing them upon the analytical process and results as much as possible.

4.8 MIXED ANALYSIS: REVISITING INTEGRATION AND SYNERGY

The analyses described in this section were designed to address RQ3, building on the discussion of integration in Section 4.3.4. Techniques reflect some of the recommendations of Bazeley (2012). It should not be overlooked, however, that mixed analysis occurs throughout the project in somewhat subtler ways, as analysis for both strands was conducted not in a strict sequence but with iteration between the two, allowing for new ideas to form in the process (Greene, 2007).

Because of the small sample size, it was not practical or useful to consider performing advanced statistical tests based on quantized qualitative data. Quantizing, in the context of this project, was done only in a few basic ways. This involved, for example, cross-tabulating specific codes against stakeholder groups or other categories (e.g. documents vs. interviews, high/middle/low performing schools, etc.) to search for patterns, and referring to the breakdown of themes within a group (a particular school, all teachers, etc.).

Using the school performance categories derived from residuals translated quantitative results into qualitative categories; these were then used to code the qualitative data. This allowed for an exploration of whether themes from the qualitative analysis help to explain or elaborate differences in effectiveness between schools or between networks according to the categories defined by qualitizing the quantitative results. This was of particular interest through the lens of the research questions, given the overall aims of the study.

Mixed methods interpretations of these analyses focus on considering how qualitative findings might help to account for and illuminate understandings of network effects (or lack thereof) and differences between schools in their effects, and discussing the implications of specific quantitative patterns associated with particular qualitative themes (relating to school improvement processes and experiences and perceptions of networks).

4.9 COMMENTS ON VALIDITY

4.9.1 Quantitative

As the quantitative analysis used secondary data, questions of construct validity apply principally to the assessments measuring literacy and math attainment. The technical reports for the assessments examine the construct validity of the items included in these assessments (CTB/McGraw Hill, 2010a; Pearson, 2012b).

Some threats to internal validity (broadly interpreted as the fit between the statistical approaches and the relevant research questions) and external validity (or generalizability of findings) were present in the research design. These validity challenges relate to issues of measurement error and missing data, both of which have been discussed in this chapter along with approaches taken to address these concerns.

4.9.2 Qualitative

Validity and reliability, constructs arising from quantitative approaches, can be reframed in the context of qualitative research as trustworthiness and dependability (Lincoln & Guba, 1985). To address these, several steps were taken during the research process. First, as recommended in case study texts, a database was kept including transcripts and researcher journal entries and memos (Stake, 1995; Yin, 2014). This is thought to aid in the process of monitoring the impact of the researcher on the research process and outcomes (Simons, 2009). This added rigour, thus enhancing the ability to provide an authentic and trustworthy account, and also addressed the dependability of the results, as careful documentation establishes a clear "chain of evidence" between the data and the interpretations and conclusions drawn (Yin, 2014). Second, multiple perspectives were taken into account. This allowed for triangulation to investigate whether findings from different stakeholder groups converged (Greene et al., 1989; Morse, 1991). Academic critiques of my work, such as from my supervisors and in response to conference presentations, also provided external checks on my interpretations.

4.9.3 Mixed

The validity of a mixed methods study must logically depend on the validity of its quantitative and qualitative strands. There are, however, additional considerations regarding validity of the mixed methods study beyond the validity of its parts. Onwuegbuzie and Johnson (2006) propose that the term "legitimation", rather than validity, be used for mixed methods studies, and offer a framework for assessing various types of legitimation. Rather than an exhaustive list of all possible types of legitimation, the points below highlight legitimation concerns of particular relevance to the present study:

1) Sample integration legitimation: This is mainly defined in terms of the relationship between qualitative and quantitative sampling, and the extent to which this allows

for quality meta-inferences through integrated analysis. In the present study, the qualitative sample was selected from within the quantitative sample, and there is a clear connection between the two strands (based on performance categories in each subject), so that there was a strong foundation for meta-inferences across strands.

2) Inside-outside legitimation: In the quantitative strand, my role was that of an observer, analysing the data objectively. In the qualitative strand, as discussed in Section 4.7.2, I took on a more blurred role somewhere between insider and outsider. Inside-outside legitimation refers to the extent to which these perspectives are reconciled to allow for justifiable inferences. In this study, the approaches used to ensure this type of legitimation included peer (and supervisor) review and feedback on reports of findings.
3) Conversion legitimation: When data is quantized or qualitized, there is an inherent risk of misrepresenting the original data. In this research, one such risk arises out of the categorisation of schools as high, typical, or low-performing; specifically, it is possible that two schools in neighbouring categories may in fact not be much different (e.g. a low-performing school just below the cut-off, and a typically-performing school just above it). To mitigate this potential problem, this issue is explicitly discussed in the reporting of findings.
4) Multiple validities legitimation: This is chiefly concerned with the validity of each strand of a mixed methods study within its own paradigm. For the current study, multiple validities legitimation is addressed by the discussion of quantitative and qualitative validity and reliability issues in Sections 4.9.1 and 4.9.2 above.

4.10 COMMENTS ON ACCESS

4.10.1 Access to data

The NYCDOE RPG was invaluable throughout the life of the study, responding to my questions about datasets and providing additional network affiliation data for which there was no formal request procedure.

Gaining access to anonymous student-level data (which is treated as confidential by NYCDOE rules) was a time-consuming process, but straightforward enough, as there are established protocols in place for data requests to the NYCDOE. Access to higher-level data was somewhat more challenging, mainly because the data stored on school, network

and NYCDOE websites are not entirely consistent nor intuitively organized. Internet archive sites were tremendously useful to mitigate some of the inconsistencies, as it was possible to find and download older documents that had been removed (versions of datasets from previous years, for example).

Network data were the most difficult to locate and interpret; some networks had publicly-accessible, updated websites with a wide variety of information available, others had websites that were over a year out of date, and some included password protection to restrict access to specific individuals or groups. Network evaluation data was available from the NYCDOE, but only for one year, and this was apparently beyond the purview of the RPG. Supporting documentation explaining the values in these data was only found through creative internet searches, which led me to a blog external to the NYCDOE on which the relevant documentation had been posted.

4.10.2 Access to individuals

One of the major challenges involved in conducting this research was that, while recruitment of schools and individuals for qualitative case studies was in progress, the topic of the project became politically sensitized because of a shift in political administration. As discussed in Chapter 2, the New York City mayor's office, and specifically a chancellor appointed by the mayor, is the highest authority over the NYCDOE, with the power to make major restructuring decisions. Perhaps because Children First Networks had been an initiative designed and spearheaded by the previous mayoral administration, more than a few participants expressed their view that the new administration elected in 2012 was likely to dissolve the network structure in favour of a more geographically based model. This may have been a particular deterrent to participation in my study by network staff, which I posit as a major reason that the number of participants successfully recruited within this stakeholder group is so small.

Gaining access to principals was straightforward to begin with, because principals' contact information is available from CEP documents and school phone numbers are posted on school websites. However, the response rate was extremely low, regardless of the method of contact. I contacted over 300 principals to gain a sample of ten, and among those ten, several asked how many others I had been able to recruit and expressed surprise that there were this many. This made the systematic sampling process in the original research design impossible, and required a reconceptualization of the case study phase of the project.

The response rate for teachers within these 10 schools was lower. One of the teachers I interviewed, after asking me whether I was getting many teacher participants, guessed that "most people just aren't going to be interested in taking time to talk about something outside their own classroom". Flyers were distributed in every school in the sample, and I made an effort to visit and introduce myself during a staff meeting at each school, provided the principal was amenable to this. Ultimately, the best response rates were in schools where the principals preferred to introduce me to teachers. This does raise concerns about selection bias, although the teachers included within each school were never intended to be a random sample. Ethical considerations also arise from this particular route to accessing teachers, and these are addressed in the section below.

4.11 ETHICAL CONSIDERATIONS

For this research, ethical approval was sought and received from both the University of Oxford Central Research Ethics Committee (CUREC) and the NYCDOE Internal Review Board (IRB) (through which researchers must apply before conducting research in New York City schools). The project was deemed to be of minimal risk to participants. Protecting participants' rights to autonomy, privacy, and respect, as emphasized in the ethical guidelines published by the British Educational Research Association (BERA, 2011).

Precautions were taken throughout and following data collection and analysis to protect the identities of participants, as well as to prevent unauthorized access to sensitive data. In the case of qualitative data, care was taken to redact or remove any identifying information from interview transcripts. Participants and organizations (schools, networks) were assigned randomized codes, with the master list kept encrypted on a hard drive separate from qualitative data. For quantitative student records, which were in any case de-identified by the NYCDOE with matched scrambled student identification numbers, any original or cleaned data were kept encrypted on drives accessed only by me, as even de-identified student data must be treated as confidential according to NYCDOE IRB rules. An essential part of the data security plan also included a timeline for destruction of the data (3 years after the completion of the project). I was required to complete a signed and notarized nondisclosure agreement to this effect, and to undertake an online human subjects research training based on the US Department of Health and Human Services guidelines (DHHS, 2009) prior to my proposal's approval by the IRB.

Voluntary and informed consent was prioritized throughout, in the interest of respecting individual participants' right to autonomy (Howe & Moses, 1999). In recruiting principals and network staff, this involved contacting each participant individually to explain the nature and purpose of the project and my role in it, making it clear that there was no obligation to participate and that participants were free to withdraw consent at any time (see Appendix 4 for the consent forms used for individuals). Recruiting teachers was somewhat more complicated. In most cases flyers (see Appendix 3) were distributed in schools to offer information about the project and its purpose, and to allow teachers to contact me directly if they wished to participate. In some schools, principals preferred to have a role in recruitment (quite possibly to prevent the possibility of my bothering teachers), and any conversations they may have had with teachers about the project were out of my control. To mitigate any potential risks associated with this, teachers in these schools were privately given the option to opt out at the beginning of the interview without my informing anyone else, including the principal, of their decision.

Brooks, Te Riele and Maguire (2014) raise the question of ethics in data analysis, specifically the concern that the researcher's values position might affect what she finds in the data. This has been addressed in the discussion of positioning in Section 4.7, and in Section 4.9.2 regarding validity of qualitative findings. Making findings accessible to several audiences (researchers, practitioners and policymakers) after the project and thesis is complete would be one way to invite critique and ensure the integrity of findings and interpretations.

Another ethical issue pertaining to the quantitative data and analyses merits discussion here, namely, the use of state assessment scores as outcomes. This makes an implicit statement about the primacy of cognitive attainment outcomes, and more specifically the limited focus on literacy and mathematics over other subjects. This statement is not one that I support in the abstract as a researcher. Rather, the use of literacy and mathematics assessment outcomes is based on two practicalities: First, these are used as major determinants of student and school success and failure in context, and thus are relevant because they are treated as important in local policy and by schools; and second, these are, to my knowledge, the only available student-level data for the entire school district at the time of this study, making it unfeasible to include alternative (e.g. socio-emotional) outcomes that may well be of at least equal interest and importance. Also, as the stated

purpose of networks was to promote student achievement, I considered it highly that school and network activity would link with this focus on academic outcomes.

These state assessments, however, are not without controversy with regard to their reliability and validity, which constitutes an additional ethical consideration. Technical reports available from the New York State Education Department offer within-year evidence of both validity and reliability (CTB/McGraw Hill, 2010a, 2010b, 2011a, 2011b, Pearson, 2012a, 2012b, 2013). However, changes from year to year (including the content of the ELA and mathematics assessments, the companies – first CTB/McGraw Hill, then Pearson – responsible for constructing and validating these assessments, and policies governing various aspects of test administration) have led to popular criticism and concern regarding the comparability of scores over time as well as the appropriateness of their use for accountability purposes (see, for example, Hernandez & Gebeloff, 2013). In fact, some of the qualitative findings presented in Chapter 6 of this thesis demonstrate that teachers and principals within the sample for the present study had concerns about both the content of these assessments and their use for the purposes of evaluating teachers and schools. Further, Ng and Koretz (2015) show that a different (and equally conventional) approach to scaling these assessments, as compared to the approach used for accountability in context, can substantially affect judgments about school quality based on either rankings and performance bands. Taking all of these issues into consideration, while New York State assessment scores were the best (and only) available measure of student achievement for the purposes of this research, using them to study school effects is arguably problematic. While it was not feasible for me to construct alternative measures, presenting the reader with a transparent account of the controversies and criticisms noted above may at least serve to mitigate ethical issues associated with the use of these assessment scores.

Finally, there is an ethical responsibility on my part to disseminate the results of the study, both to policymakers in the relevant context and to the research community more broadly. For policymakers, this is important because findings from this study may be able to inform relevant research-based decisions for similar structures and practices in the future. This is especially pertinent in a system that has seen so much and such rapid change with regard to a wide variety of policies and restructuring. There is an advocacy element to making results available to practitioners, in that doing so has the potential to spark some interest and investment on the part of teachers and principals, which may encourage them to examine and to be proactive about their needs and views relevant to the topic at hand.

Disseminating results to the wider research community through publication will also be necessary, both to allow for criticism and critique, and to make a contribution to the relevant knowledge base. To this end, some of the results and methodological aspects of this research have been presented at a number of international conferences at various points during the project (Lindorff-Vijayendran, 2014, 2015a, 2015b, 2015c).

4.12 LIMITATIONS OF THE STUDY

Several limitations of the study have already been addressed in various sections throughout this chapter. These are summarised below.

First, the quantitative analysis is limited by the quality and type of data available. Some variables likely to be relevant based on established and tested models of educational effectiveness were not available, so that the predictors included in multilevel models were somewhat opportunistic by necessity. Additionally, because test scores are not vertically equated, growth curve analysis was not feasible. Such an analysis might have presented a different picture of network, district and school effects on achievement growth over several time points, as previous research suggests that schools' effects on achievement growth over time can be considerably larger than effects on achievement status at a particular time (Teddlie & Reynolds, 2000). This also imposes limitations on the study's ability to make inferences about the relationship(s) between network affiliation and attainment over time.

Another limitation of the study concerns the (potential lack of) immediate practical utility of findings, and stems from the rapid and on-going change in politics and policy in the district. The network structure that formed the focus of this research has only been in place since 2010-11, but was already being considered for restructuring by the time data collection was complete. This may have affected how some participants reacted to the research focus as well as their willingness to participate, and some explicitly expressed that they were sceptical about the practical value of the research because they did not expect the current structure or policy to remain in place for long. It is my assertion, however, that findings from this study can still prove valuable as a contribution to the broader literature on school networks and collaboration.

The timing of data collection presented an additional limitation. Not all participants were interviewed at even approximately the same time; in fact, though data collection was

planned for the Spring semester of 2014, some participants did not respond to recruitment emails or calls until several months later, which meant that interviews were not completed until January 2015. This was taken into account in analysing the interview data, specifically by looking for any patterns in responses from different school years (2013-14 versus 2014-15) and terms (Spring and Fall). Nonetheless, there may have been some issues around retrospective perspectives, as later interviewees had to remember further back to speak about, for example, school culture prior to the most recent network affiliation.

Results from the case study must be interpreted cautiously because of the small qualitative sample size, and because of the hurdles to systematic purposive sampling of schools described in Section 4.6.2. Because not every school's participation was identical, this made comparisons across networks and schools challenging. The mechanisms of any potential bias in the school sampling process are not immediately obvious from a preliminary analysis of findings, but this does not preclude the possibility of more nuanced response bias. The fact that it was not possible to recruit several schools in the same network was also problematic. By reframing the case study to focus on the network structure as a whole, and the experience of it in schools, this issue was at least partially resolved, as this avoids unsupported comparisons between unobserved or under-observed sub-units of analysis (individual networks). Setting aside notions of comparison between cases, analytical generalisations can reasonably be undertaken, from case to theory rather than from case to some other similar case (Yin, 2014). The intention was to create a plausible and trustworthy account to enhance understanding of the ways in which participants experienced networks in their professional work using the embedded case study approach.

In order to preserve the integrity of the research and to avoid over-claiming or over-generalising, the chapters that follow explicitly reference these limitations in the context of the findings reported in order to allow the reader to interpret results appropriately.

5 EXPLORING AND EXPLAINING ACHIEVEMENT VARIATION IN THE NETWORK STRUCTURE FOR SCHOOL SUPPORT

5.1 CHAPTER GUIDE

This chapter is intended to:

- Link the contents of this chapter to the overall study aims
- Identify the research questions addressed in the chapter, and explain the approaches taken to answer them
- Present and interpret findings from each stage of analysis
- Relate analytical decisions and results to the relevant methodological and substantive literature
- Synthesize key findings and discuss their implications
- Demonstrate how results from this chapter inform subsequent data collection and analysis

5.2 CHAPTER INTRODUCTION

Based on the claims quoted in Chapter 1 (Introduction), the stated purpose of the networks on which this research focuses is to provide differentiated support for schools, ultimately raising student achievement. The latter point is underscored by the fact that networks are "held accountable for both principal satisfaction and the academic performance of the schools they support" (NYCDOE, 2013, p.1). Additionally, because schools may choose or change networks annually, there is an implicit assumption that there are meaningful differences between networks, either in terms of student or school performance, or in terms of the supports and services offered. The primary goal of this chapter, as well as an overall aim of this research, is to investigate whether there is empirical support for the claims made about networks and student achievement in context.

In addition to address a major aim of the study and a specific subset of research questions (see Section 5.3 below), the results presented in this chapter also serve to explain how relative school effectiveness was defined and used as a basis for the selection and recruitment of schools for qualitative case study.

5.3 APPROACH TO ANALYSIS

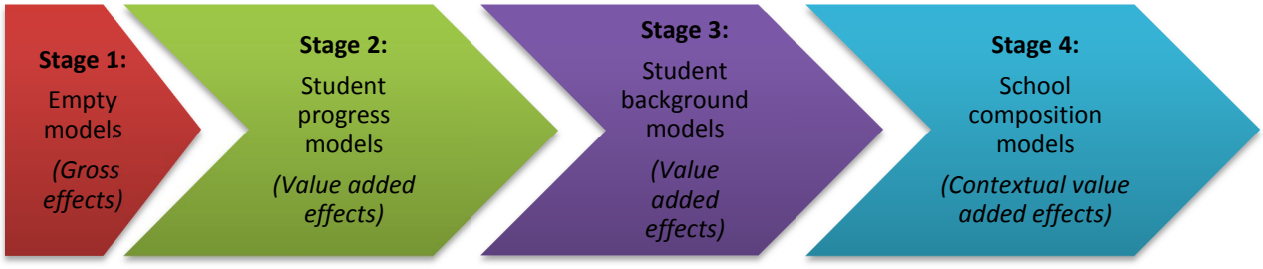
Results presented in this chapter address the following subset of the research questions from Chapter 4 (Section 4.2):

RQ1: *Is there evidence to suggest that networks have measurable effects on student and school achievement?*

- ***RQ1a:*** *How does 8th grade achievement in literacy and mathematics vary between networks, districts, schools, and students, and is there evidence of significant overall effects at these levels?*
- ***RQ1b:*** *To what extent can variation in 8th grade achievement be statistically explained at each level by student background factors, prior attainment in 5th grade, school context and composition, and network or network-affiliation characteristics (e.g. duration of affiliation)?*
- ***RQ1c****: How consistent are apparent effects within years across subjects (literacy and mathematics), and how stable are they within subjects across years?*

As explained in the description of quantitative methods (Section 4.5.4), multilevel modeling was selected as the approach best suited to answering these research questions. Results are based on random coefficient models, which extend traditional multiple regression techniques by allowing coefficients to vary for units at levels above individual observations (in the context of this research, these higher levels include schools, networks, and districts, with individual observations at the student level). Analysis proceeds through four stages, outlined in Figure 7 and described in detail below.

Figure 7 Stages of analysis: Gross, value added and contextualised value added models



First, to address **RQ1a**, variance was partitioned between the student, school, network and/or district levels in four model specifications each for literacy and mathematics achievement outcomes. This approach facilitated the calculation of overall "gross" effects

(Scheerens & Bosker, 1997) at each level based on the proportions of total variance lying between schools, between networks, and between districts, without correcting for any additional variables. Empty (unadjusted) models were compared for each subject to determine whether each level (or classification, in the case of cross-classified models with districts and networks both at the third level) could be interpreted as having an overall effect on student achievement. Comparisons of model fit at this stage thus informed the selection of models built upon in subsequent analyses.

Following these initial comparisons, the second stage of analysis involved adding predictors to control for students' individual prior achievement. The importance of controlling for measures of prior ability is well established in the educational effectiveness literature (Aitkin & Longford, 1986; Ballou, Sanders, & Wright, 2004; Goldstein, 1997; P. W. Hill & Rowe, 1998; Opdenakker & Van Damme, 2001; Sammons et al., 1993; Thomas, 1995), as this allows for interpretations based on schools' contributions to student progress rather than merely their association with students' final achievement outcomes.

The third stage of model fitting introduced additional predictors at the student level to account for individuals' background characteristics, such as gender, ethnicity, free or reduced school lunch eligibility as a proxy for socioeconomic status (SES), English language proficiency, special educational needs, and age. Previous research has underscored the importance of accounting for available student background variables, as these often constitute significant predictors of achievement and account for a substantial proportion of between-student variance (Lee & Bryk, 1989; Raudenbush & Bryk, 1986; Thomas & Mortimore, 1996; Willms, 1992). Ultimately, controlling for student background is expected to influence the calculation of school effects (Sammons, 1999).

Finally, the random coefficient models were extended to include variables relevant to school composition and context. While not all studies of educational effectiveness control for such predictors, there is substantial empirical evidence to suggest that controlling for school composition and context often has considerable influence on the rankings and magnitudes of school effects (Palardy, 2008; Timmermans & Thomas, 2014; Willms, 2010; Willms & Raudenbush, 1989). By accounting for school composition, the goal is ultimately to separate out the effects of aspects of the school context and population that are out of the school's control; in other words, the intention is to be able to calculate relative school effectiveness based on school processes and interactions after controlling

for characteristics of the student population and the broader context in which schools operate. In the present study, some of the predictors tested for inclusion at this stage were proportions of particular categories of students (for example, proportions of Black and Hispanic students and students eligible for free or reduced school lunch), while others related to grades served (K-8, 6-8 only, or K-12) and total enrolment. Additional variables considered for inclusion at this stage related to schools' network affiliation, including length of time affiliated with a particular network, network mix of school levels (numbers of elementary, middle, and high schools), overall network size (based on the total number of schools), as well as an indicator of whether a network was linked to an external Partnership Support Organization or entirely internal to the NYCDOE.

At each stage of analysis, there were several key considerations. These include the following, with the last three items applying only to models fitted after the first stage:

1. Model fit: Because Markov Chain Monte Carlo (MCMC) estimation is used, the appropriate indicator of model fit is the Deviance Information Criterion (DIC) developed by Spiegelhalter et al. (2002), which penalises for the effective number of parameters included in the model. As is true for all statistical tests, the cut-off values for determining significance are somewhat arbitrary. While recommendations vary between authors, this study adopts the conservative rule of thumb that a difference of 7 or more units provides strong support for the selection of the model with a lower DIC value (Ibid). The DIC has the additional benefit of allowing for comparisons between non-nested models, which is especially important in the first stage of analysis to compare models including different combinations of levels and cross-classifications. Along with DIC values, it is also important to consider whether MCMC chains for the parameters in each model converge. Even if a model demonstrates a substantially reduced DIC value, inferences about a parameter are not supported if its MCMC chain shows bad mixing that cannot be improved by increasing the burn-in and total chain lengths (this technical issue is of particular importance in comparing empty models, as the addition of fixed parameters did not lead to significant changes in convergence).
2. Intra-class correlation (ICC): The ICC (ρ) for a particular level of a model can be interpreted in two ways. In empty models, the ICC is the proportion of variance that lies between units at a particular level. For models with additional predictors there are multiple approaches to calculating the ICC, either as a correlation

between two students in the same school, or as the proportion of residual unexplained variance lying between units at a given level. Results given in this chapter adhere to the latter approach, using a for residual ICC (ρ) Raudenbush and Bryk (2002),

$$\rho = \frac{\sigma_{u0}^2}{\sigma_{u0}^2 + \sigma_e^2},$$

where σ_{u0}^2 and σ_e^2 represent the variances at the school and student levels, respectively.

3. Consistency: The results presented for each model included in this chapter separately consider 8th grade student achievement in two subjects, literacy and mathematics, in 2011-12. It was therefore useful to compare the results obtained separately for each subject to assess the extent to which school effects are consistent in literacy and mathematics (i.e. the extent to which schools that are effective in literacy also tend to be effective in mathematics, and vice versa). Correlations between the school-level residuals obtained from models fitted to literacy and mathematics data from 2011-12 provided a way to measure this consistency. These correlations were weighted by the inverse of the squared average standard error of each pair of residuals compared (based on Hedges' proposal of inverse variance weighting, e.g. in Hedges & Olkin, 1985); failing to weight such calculations would fail to account for the different numbers of students per school and the related standard errors of school effects, which would risk misleading interpretations of consistency. An identical procedure was used to compare school effects over different years to assess the stability of school effects over time in Chapter 7. These comparisons collectively address **RQ1c**; the present chapter focuses on the first part of the question addressing consistency across subjects because it focuses on models for data from a particular year used to inform sampling, with these models extended to address the question of stability in Chapter 7. An alternative approach might have involved estimating multivariate response models with literacy and mathematics scores nested within students. However, MLwiN was not able to handle models with the number of classifications required to estimate such models for the first set of model comparisons, particularly the cross-classified specification in which schools were nested within districts and

also within networks. Additionally, even for the two-level models, multivariate response specifications required considerably longer to run using MCMC estimation in MLwiN (up to 8 times the amount of time required for single-response specifications, or more than a day and a half for each individual estimation). Taking into consideration the cost in terms of computation time versus the information gained by running multivariate response models, it seemed far more efficient and practical to estimate separate models for each subject and then calculate correlations between the two at the level of interest (i.e. at the school level).

4. Variance explained: In each model in which new predictors are included, the amount of variance explained at each level is calculated as a proportion of the original variance at that level in the corresponding empty model. This calculation informs interpretations of the influence of various groups of predictors (i.e. student prior attainment, student background, and school context and composition) at each level of the model of interest, addressing **RQ1b.**
5. Significance of individual predictors: As is usual in single-level regression analyses, a common method for testing the significance of coefficients in multilevel models is to use t-tests for fixed coefficients, and Wald z-tests and likelihood ratio tests for random parameters. However, MCMC estimation is an approach based on Bayesian statistics, and in a Bayesian framework hypothesis testing is understood somewhat differently to that of frequentist statistical methods. For results obtained via MCMC estimation, Bayesian credible intervals for fixed coefficients and differences in DIC for random parameters can be used to assess significance. The use of DIC values is as described above, while for each fixed coefficients, the crucial consideration in Bayesian significance testing is whether its corresponding credible interval contains zero (if so, the parameter is not judged to be significantly different from zero). The interpretation of a Bayesian credible interval is quite straightforward: Based on a specific set of data, a 95% credible interval for a specific parameter implies a 95% probability that the parameter lies within this interval. For the results reported in this thesis, the significance of fixed and random parameters was checked using both frequentist and Bayesian methods at various stages of analysis to ensure that inferences were similar regardless of the approach used. Both pooled standard errors and credible intervals are reported in tabulated model results on the basis that standard errors provide a clear measure of

total variability (between and within imputations) for results obtained after multilevel multiple imputation of missing data, and credible intervals provide an appropriate analogue for significance testing within the framework of Bayesian statistics.

While models were fitted separately for 8th grade assessment outcomes in four separate years in this study (2009-10, 2010-11, 2011-12, and 2012-13), the body of this chapter includes full results only from the 2011-12 models. This was done to simplify the presentation of results for the reader, as well as because it was the 2011-12 results that were used directly to inform sampling for the qualitative strand of the study.

5.3.1 Model specifications

The sections that follow present equations corresponding to the four progressive stages of analysis described above. Hox's (2010) recommendation regarding notation is followed, that is, β is used to represent regression coefficients and u is used for residual variance terms, with subscripts to indicate the levels over which each term varies. Operational definitions of variables are as given in the previous chapter (Section 4.5.1).

5.3.1.1 Empty models

Equations 5.1 through 5.4 specify the empty models considered at the first stage of analysis. In equations 5.1 through 5.3, individuals (i) are nested within schools (j), nested (as applicable) in networks (k) or districts (l).

Equation 5-1 2-level empty model: Students within schools

$$y_{ij} = \beta_0 + u_{0j} + e_{ij}$$

$$u_{0j} \sim N\left(0, \sigma^2_{u_j}\right)$$

$$e_{ij} \sim N(0, \sigma^2_e)$$

Equation 5-2 3-level empty model: Students within schools within networks

$$y_{ij} = \beta_0 + u_{0k} + u_{0jk} + e_{0ijk}$$

$$u_{0k} \sim N\left(0, \sigma^2_{u_k}\right)$$

$$u_{0jk} \sim N\left(0, \sigma^2_{u_j}\right)$$

$$e_{0ijk} \sim N(0, \sigma^2_e)$$

Equation 5-3 3-level empty model: Students within schools within districts

$$y_{ij} = \beta_0 + u_{0l} + u_{0jl} + e_{0ijl}$$

$$u_{0l} \sim N\left(0, \sigma^2_{u_l}\right)$$

$$u_{0jl} \sim N\left(0, \sigma^2_{u_j}\right)$$

$$e_{0ijl} \sim N(0, \sigma^2_e)$$

In each equation above, the dependent variable (y) represents the student normalised literacy or mathematics score, with subscripts to indicate the levels included in the model. At this stage, the fixed part of each model includes only a random intercept term (β_0), again with additional subscripts to identify the levels across which this intercept is allowed to vary. The random part of each model includes the residual terms for each level included in each model $(u_{0l}, u_{0k}, u_{0j}, e_{0ij})$. Distributions of residuals at each level are assumed to be normal, and their variances $(\sigma^2_{u_l},\ \sigma^2_{u_k},\ \sigma^2_{u_j},\ \sigma^2_e)$ homoscedastic; these assumptions are checked following the presentation of results.

Different notation is needed to specify the cross-classified models; levels are no longer strictly nested, so that the previous system of subscripts would be potentially misleading. Instead, a slight adaptation of the classification notation proposed by Browne, Goldstein, and Rasbash (2001) is used below.

Equation 5-4 4-classification empty model: Students within schools, cross-classified within networks and districts

$$y_i = \beta_0 + u^{(4)}_{0,district(school(i))} + u^{(3)}_{0,network(school(i))} + u^{(2)}_{0,school(i)} + e_i$$

$$u^{(4)}_{0,district(school(i))} \sim N\left(0, \Omega^{(4)}_{u0}\right)$$

$$u^{(3)}_{0,network(school(i))} \sim N\left(0, \Omega^{(3)}_{u0}\right)$$

$$u^{(2)}_{0,school(i)} \sim N\left(0, \Omega^{(2)}_{u0}\right)$$

$$e_i \sim N\left(0, \Omega_e\right)$$

In the above model specification, $u^{(2)}_{0,school(i)}$ and e_i are equivalent to school and student residual terms u_{0j} and e_{ij} in previous models. Superscripted classifications distinguish between the higher levels of the model. Subscripts for higher-level terms represent the

partial nesting in the model, i.e. students within schools (*school(i)*), schools within districts (*district(school(i))*), and schools also within networks (*network(school(i))*). Variance at each level is more generally represented ($\Omega_{u0}^{(4)}, \Omega_{u0}^{(3)}, \Omega_{u0}^{(2)}, \Omega_e$) to accommodate the superscripted classifications.

5.3.1.2 Prior attainment value added models

As noted in Section 5.3, in order to model student achievement and school effects, it is recognised as important to control for prior attainment to arrive at "value added" models. Using notation as above for equations without cross-classification, the second set of models are specified by Equations 5.5 and 5.6.

Equation 5-5 2-level prior attainment value added model (literacy)

$$y_{ij} = \beta_0 + \beta_1\left(PRILIT_{ij}\right) + \beta_2\left(PRILIT_{ij}\right)^2 + \beta_3\left(PRIMAT_{ij}\right) + u_{0j} + e_{ij}$$

$$u_{0j} \sim N(0, \sigma_{u0}^2)$$

$$e_{ij} \sim N(0, \sigma_e^2)$$

Equation 5-6 2-level prior attainment value added model (mathematics)

$$y_{ij} = \beta_0 + \beta_1\left(PRIMAT_{ij}\right) + \beta_2\left(PRIMAT_{ij}\right)^2 + \beta_3\left(PRILIT_{ij}\right) + u_{0j} + e_{ij}$$

$$u_{0j} \sim N(0, \sigma_{u0}^2)$$

$$e_{ij} \sim N(0, \sigma_e^2)$$

The additional coefficients in these equations correspond to linear (β_1) and quadratic (β_2) coefficients for normalised prior achievement in the model subject of interest (e.g. prior literacy achievement in the model with literacy achievement as the dependent variable), and a coefficient for normalised prior achievement in the alternative subject (e.g. prior mathematics achievement in the model with literacy achievement as the dependent variable, denoted by (β_3). Besides presenting results for these models, Section 5.6.1 also shows how they were built up from simpler models without the quadratic or alternative subject prior achievement terms.

5.3.1.3 Student background value added models

In order to contextualise the models, EER on school effects typically controls for a range of potential background influences, such as student characteristics, to further refine value added models and distinguish the effects of schools from students' contexts. To accomplish this, additional predictors are introduced into the literacy and mathematics models at this stage, shown in Equations 5.7 and 5.8.

Equation 5-7 2-level prior attainment and student background value added model (mathematics

$$y_{ij} = \beta_0 + \beta_1(PRILIT_{ij}) + \beta_2(PRILIT_{ij})^2 + \beta_3(PRIMAT_{ij}) + \beta_4(AGE_{ij}) + \beta_5(FSM_{ij}) + \beta_6(LANGPROF_{ij}) + \beta_7(ETH{:}BLACK_{ij}) + \beta_8(ETH{:}HISPANIC_{ij}) + \beta_9(ETH{:}ASIAN_{ij}) + +\beta_{10}(ETH{:}NATIVE\ AMERICAN_{ij}) + \beta_{11}(ETH{:}MULTI_{ij}) + \beta_{12}(MALE_{ij}) + \beta_{13}(SWD_{ij}) + u_{0j} + e_{ij}$$

$$u_{0j} \sim N(0, \sigma_{u0}^2)$$

$$e_{ij} \sim N(0, \sigma_e^2)$$

Equation 5-8 2-level prior attainment and student background value added model (mathematics)

$$y_{ij} = \beta_0 + \beta_1(PRIMAT_{ij}) + \beta_2(PRIMAT_{ij})^2 + \beta_3(PRILIT_{ij}) + \beta_4(AGE_{ij}) + \beta_5(FSM_{ij}) + \beta_6(LANGPROF_{ij}) + \beta_7(ETH{:}BLACK_{ij}) + \beta_8(ETH{:}HISPANIC_{ij}) + \beta_9(ETH{:}ASIAN_{ij}) + +\beta_{10}(ETH{:}NATIVE\ AMERICAN_{ij}) + \beta_{11}(ETH{:}MULTI_{ij}) + \beta_{12}(MALE_{ij}) + \beta_{13}(SWD_{ij}) + u_{0j} + e_{ij}$$

$$u_{0j} \sim N(0, \sigma_{u0}^2)$$

$$e_{ij} \sim N(0, \sigma_e^2)$$

Newly-introduced coefficients at this stage correspond to student-level variables for age (β_4), eligibility for free/reduced lunch (β_5), English language proficiency (β_6), ethnicity (β_7 through β_{11}), gender (β_{12}), and disability (β_{13}).

5.3.1.4 School composition and context value added models

The final models presented in this chapter further contextualise results by controlling for variation in school composition. These are specified by Equations (5.9) and (5.10) below.

Equation 5-9 2-level contextual value added model (literacy)

$$y_{ij} = \beta_0 + \beta_1(PRILIT_{ij}) + \beta_2(PRILIT_{ij})^2 + \beta_3(PRIMAT_{ij}) + \beta_4(AGE_{ij}) + \beta_5(FSM_{ij}) + \beta_6(LANGPROF_{ij}) + \beta_7(ETH{:}\,BLACK_{ij}) + \beta_8(ETH{:}\,HISPANIC_{ij}) + \beta_9(ETH{:}\,ASIAN_{ij}) + +\beta_{10}(ETH{:}\,NATIVE\ AMERICAN_{ij}) + \beta_{11}(ETH{:}\,MULTI_{ij}) + \beta_{12}(MALE_{ij}) + \beta_{13}(SWD_{ij}) + \beta_{14}(SCHPRILIT_j) + u_{0j} + e_{ij}$$

$$u_{0j} \sim N(0, \sigma_{u0}^2)$$

$$e_{ij} \sim N(0, \sigma_e^2)$$

Equation 5-10 2-level contextual value added model (mathematics)

$$y_{ij} = \beta_0 + \beta_1(PRIMAT_{ij}) + \beta_2(PRIMAT_{ij})^2 + \beta_3(PRILIT_{ij}) + \beta_4(AGE_{ij}) + \beta_5(FSM_{ij}) + \beta_6(LANGPROF_{ij}) + \beta_7(ETH{:}\,BLACK_{ij}) + \beta_8(ETH{:}\,HISPANIC_{ij}) + \beta_9(ETH{:}\,ASIAN_{ij}) + +\beta_{10}(ETH{:}\,NATIVE\ AMERICAN_{ij}) + \beta_{11}(ETH{:}\,MULTI_{ij}) + \beta_{12}(MALE_{ij}) + \beta_{13}(SWD_{ij}) + \beta_{14}(SCHPRIMAT) + u_{0j} + e_{ij}$$

$$u_{0j} \sim N(0, \sigma_{u0}^2)$$

$$e_{ij} \sim N(0, \sigma_e^2)$$

In the literacy model, only one new terms is introduced. This (β_{14}) represents the effect of school mean prior (5th grade) achievement in literacy (centred on the citywide average for the corresponding year); other school- and network-level variables were tested and rejected for inclusion in the model. A corresponding term is included in the mathematics model, representing the effect of school mean prior achievement in mathematics, again centred on the citywide mean prior achievement value.

5.4 DESCRIPTION OF SAMPLE DATA

Univariate statistics for individual variables computed over the entire population of 8th grade students in 2009-10, 2010-11 and 2011-12 were given in the previous chapter (Section 4.5.2). Tables 7 and 8 offer a more detailed description of 8th grade scale scores in literacy and mathematics (unstandardized, to contrast actual point values on the original

scale), broken down by categories of student characteristics. These results provide some preliminary insight into the relative performance of various groups of students by gender, ethnicity, SES, English language proficiency, disability status, age, and prior achievement.

The findings suggest that female students appear to outperform male students in both literacy and mathematics, while the scores of male students in literacy are slightly more variable than those of their female counterparts. The total number of students in the Native American and Multi-ethnic categories of ethnicity are so low, with each making up less than less than 1% of the total sample in each subject, that comparisons between these and other ethnic groups may not be particularly meaningful. This also suggested that coefficients associated with these categories might be unlikely to be significant in subsequent analyses, as standard errors are generally inflated for small counts (and, equivalently, the corresponding Bayesian credible intervals are wide), although results show that this did not prove to be the case on a consistent basis. Amongst the ethnic categories with greater numbers of students, White students outperformed the other groups (excluding the Multi-racial category) in literacy, while Asian students had an average score that was only slightly lower but more variable. Black and Hispanic students scored similarly and lower than other groups on average, and in literacy in particular these groups demonstrated less variability than other ethnic categories.

Although it is clear and not especially surprising that students eligible for free or reduced lunch (as a proxy for SES) have lower average achievement and very slightly less variability in both subjects, this is based on a dichotomous variable (eligible/ineligible) which is at best only a very crude measure of SES. The fact that over 80% of students are designated as "eligible" in this large sample of students in New York City schools compounds the problem, so that statistical models are unlikely to pick up much of the actual variation between students based on SES. Students classified as ELLs and students with disabilities (i.e. special educational needs) demonstrated lower average achievement than students without these classifications, but had higher variability in their scores with the exception of students with disabilities in literacy. This, too, may well be because these dichotomous measures are fairly crude, and may cover a wide range of different circumstances (a recently-immigrated student who speaks Spanish and whose parents had a high level of education in her home country, for example, would be in the same category as a student with an interrupted education and speaking a language less prevalent in New

York City, despite the fact that there would likely be quite different support needs as well as different availability of appropriate support for these students).

According to these descriptive statistics, there does not seem to be much difference in average attainment in either subject between underage students, defined as being more than one standard deviation below the average age at the time of assessment, and their peers within one standard deviation of the mean age. On the other hand, overage students (more than one standard deviation above the mean age) have substantially lower achievement in both literacy and mathematics, and slightly higher variability. This likely reflects complex issues related to grade retention, and is discussed in more depth in Section 5.6.2, after controlling for age in literacy and mathematics models.

Categories of low, typical and high prior achievement in each subject were defined as more than 1 standard deviation below the mean, within one standard deviation of the mean, and more than one standard deviation above the mean, respectively. Comparing the descriptive statistics for these categories in each table yields unsurprising results regarding average achievement, students with higher scores on 5th grade assessments in either literacy or mathematics have higher achievement in their 8th grade scores in both subjects. Students with prior scores within the typical range (-1SD to 1SD) on 5th grade assessments demonstrate the least variability, which in part reflects the fact that these make up the largest proportion of the total population. The highest-achieving groups in both subjects in 5th grade also demonstrates the greatest variability in 8th grade literacy, while the same is true of the lowest-achieving groups in both subjects with regard to 8th grade mathematics.

These basic findings do not shed light on whether these specific comparisons are statistically significant, but the differences between groups of students discussed in this section provided an initial empirical rationale for considering these variables for inclusion in the more sophisticated analyses to follow. The relative size of some of these categorical groups also suggests caution in interpreting effects associated with the proxy measure of SES (due to the limitations of the indicator available), and with the two proportionally smallest ethnic categories.

Table 7 Descriptive statistics by category (Literacy, 2011-12)

		Literacy 2012			
		N	**% of sample**	**Mean**	**SD**
Gender	*Male*	34367	50.52	651.29	20.42
	Female	33656	49.48	656.82	20.36
Ethnicity	*White*	9667	14.21	663.57	20.35
	Black	19615	28.84	650.6	17.32
	Hispanic	27276	40.10	649.76	19.41
	Asian	10973	16.13	662.24	23.45
	Multi	130	0.19	669.12	22.43
	Native American	362	0.53	651.89	21.28
Free/reduced lunch	*eligible*	56798	83.50	651.98	19.98
	ineligible	11225	16.50	664.35	20.45
ELL status	*ELL*	8667	12.74	633.74	21.6
	proficient	59356	87.26	656.98	18.65
Disability	*classified*	10956	16.11	637.72	18.31
	unclassified	57067	83.89	657.15	19.49
Age	*>1SD below mean*	8591	12.63	655.21	19.76
	within 1SD of mean	50660	74.47	655.93	20.06
	>1SD above mean	8772	12.90	641.83	21.19
Prior literacy	*>1SD below mean*	6325	9.30	634.46	20.89
	within 1SD of mean	54179	79.65	654.15	17.93
	>1SD above mean	7519	11.05	669.59	24.28
Prior math	*>1SD below mean*	9170	13.48	637.63	19.23
	within 1SD of mean	47999	70.56	653.71	17.84
	>1SD above mean	10854	15.96	669.25	21.71

Table 8 Descriptive statistics by category (Mathematics, 2011-12)

		Mathematics 2011-12			
		N	% of sample	Mean	SD
Gender	*Male*	34367	50.52	677.17	30.66
	Female	33656	49.48	679.54	29.85
Ethnicity	*White*	9667	14.21	689.51	27.68
	Black	19615	28.84	669.23	27.44
	Hispanic	27276	40.10	671.8	27.73
	Asian	10973	16.13	700.98	28.92
	Multi	130	0.19	691.52	28.54
	Native American	362	0.53	676.31	30.69
FSM	*eligible*	56798	83.50	676.08	29.85
	ineligible	11225	16.50	689.78	29.93
English Proficiency	*ELL*	8667	12.74	663.09	32.64
	proficient	59356	87.26	680.57	29.28
Disability	*classified*	10956	16.11	654.64	30.09
	unclassified	57067	83.89	682.89	28.13
Age	*>1SD below mean*	8591	12.63	681.05	29.57
	within 1SD of mean	50660	74.47	680.82	29.45
	>1SD above mean	8772	12.90	661.4	30.34
Prior literacy	*>1SD below mean*	6325	9.30	656.78	34.23
	within 1SD of mean	54179	79.65	677.97	27.55
	>1SD above mean	7519	11.05	699.16	31.82
Prior math	*>1SD below mean*	9170	13.48	652.58	30.69
	within 1SD of mean	47999	70.56	677.36	25.44
	>1SD above mean	10854	15.96	704.44	29.12

5.5 PARTITIONING VARIANCE

As specified above in Section 5.3, four possible models were considered and fitted to each year's data in literacy and mathematics. In each table, models are identified as follows (followed by "a" and "b" for literacy and mathematics models, respectively):

- A-I = 2 level model, students within schools
- A-II = 3-level model, students within schools within districts
- A-III= 3-level model, students within schools within networks
- A-IV = 4-classification model, students within schools cross-classified in networks and districts

Each model was initially run for each imputed data set with a burn-in length (defined as an initial number of iterations run and discarded at the beginning of an MCMC estimation process) of 10,000 iterations, and a monitoring chain length (i.e. the number of random draws included in posterior distributions) of 20,000 iterations. For Models A-Ia and A-Ib, the diagnostics provided in MLwiN suggested that these were sufficient to reach convergence. For all other models, longer burn-in and chain lengths were indicated; these models were then re-run with burn-in and monitoring chain lengths of 100,000 and 200,000 iterations, making them quite computationally demanding. Default priors were used in the absence of appropriate informative prior values, and Iterative Generalized Least Squares (IGLS) estimates for each model were used as starting values, with the exception of the cross-classified models (IGLS estimation would have treated network and district levels as nested and produced relatively meaningless starting values, so for these models starting values were based on IGLS estimates from Models A-Ia and A-Ib).

Turning first to the literacy results in Table 9, the DIC values for Models A-Ia (178329.6), A-IIa (178326.1), A-IIIa (178328.4), and A-IVa (178326.0) are all within approximately 4 units of each other. This does not provide strong support for the selection of one model over another. The mathematics results in Table 10 provide even less support for the selection of a particular model, as Models A-Ib (175126.6), A-IIb (175124.9), A-IIIb (175126.3), and A-IVb (175124.9) differ in their DIC values by fewer than 2 units. Without further examination, these results would indicate that all four models might merit consideration in subsequent analyses. However, an inspection of the MCMC chains showed that the random parameters associated with the network and district levels have convergence problems, evidenced by bad mixing and persistent autocorrelation. Long burn-in and chain lengths (5,000,000 and 10,000,000 iterations, respectively) were run for each of these models to assess whether convergence could be improved, but there was no improvement in the mixing or autocorrelation for these parameters. Alternative techniques with the potential to improve convergence are available in MLwiN, including parameter expansion and hierarchical centring (Browne, 2015); these were explored but did not produce improved results. Thinning (which involves storing only a certain frequency of the random draws in an MCMC chain, e.g. discarding the first nine of each ten iterations) did substantially improve mixing when thinning values were set at 20 or above, but this approach has been shown to reduce estimation accuracy (MacEachern & Berliner, 1994) and so was rejected. Taking into account the long run times necessary to estimate Models

A-II, A-III and A-IV for literacy and mathematics outcomes, the convergence problems in these models, and the fairly small differences in DIC values between all models considered, Models A-Ia and A-Ib were selected as the most efficient, parsimonious and reliable choices for further analysis.

While inferences drawn from models that have failed to converge are problematic, ICCs (which are equivalent to variance components in the empty models) for all of the empty models are discussed briefly below. Because the parameter estimates and credible intervals were not much changed by running high numbers of additional iterations, these values are at least marginally meaningful (however, inferences based on these models would become increasingly unreliable with the inclusion of additional predictors).

5.5.1 Network and school effects

Network ICCs in Models A-IIIa (ρ =0.034)) and A-IIIb (ρ=0.033) suggest that approximately 3.4% of the total variation lies between networks in literacy, and 3.3% in mathematics. In the same models, school ICCs in literacy (ρ=0.179) and mathematics (ρ=0.215) show that a much greater proportion of the total variation in student achievement lies between schools for both subjects (17.9%, 21.5%), with the rest of the variation (78.6%, 75.2%) lying between students.

5.5.2 District and school effects

In Models A-IIa and A-IIb, district and school ICCs indicate that 6% of the total variation in literacy achievement lies between districts and 15.7% lies between schools, while in mathematics 7.7% of the total variation lies between districts and 17.8% lies between schools, with the remaining proportions of total variation lying between students (78.4%[11], 74.5%). Comparing these gross district effects to network effects in Models A-IIIa and A-IIIb, it appears that there was more of an association between the district to which a school belonged and achievement in both literacy and mathematics than there was between a school's network affiliation and the same set of achievement outcomes. At this stage, however, it is worth keeping in mind that as districts were geographically bounded, their apparent effects may have reflected the effects of demographic composition features of

[11] Any discrepancies of 0.01 in the sum of percents listed for a given subject are the result of rounded values for ICCs.

districts rather than of policies and practices of the offices responsible for overseeing schools at the district level.

Table 9 Empty model results, Literacy 2011-12

Literacy 2011-12												
Model	**A-IVa**			**A-IIIa**			**A-IIa**			**A-Ia**		
	Mean	S.E.	95%CI	Mean	S.E.	95%CI	Mean	S.E.	95%CI	Mean	S.E.	95%CI
Fixed part												
Intercept	-0.043	0.049	(-0.138,0.056)	-0.051	0.034	(-0.118,0.014)	-0.044	0.046	(-0.134,0.047)	-0.063	0.021	(-0.105,-0.021)
Random part												
Level: network	0.008	0.006	(0.001,0.025)	0.035	0.013	(0.014,0.067)						
Level: district	0.061	0.021	(0.03,0.112)				0.061	0.020	(0.031,0.11)			
Level: school	0.153	0.012	(0.132,0.178)	0.182	0.013	(0.158,0.211)	0.160	0.012	(0.139,0.184)	0.213	0.014	(0.187,0.243)
Level: student	0.800	0.004	(0.792,0.809)	0.800	0.004	(0.792,0.809)	0.800	0.004	(0.792,0.809)	0.800	0.004	(0.792,0.809)
DIC:	178326.0			178328.4			178326.1			178329.6		
pD:	452.0			455.5			452.1			457.2		
Units: network	57			57								
Units: district	32						32					
Units: school	477			477			477			477		
Units: student	68023			68023			68023			68023		
ρ (network ICC)	0.008			0.034								
ρ (district ICC)	0.060						0.060					
ρ (school ICC)	0.150			0.179			0.157			0.210		
ρ (student ICC)	0.782			0.786			0.784			0.790		

Table 10 Empty model results, Mathematics 2011-12

Mathematics 2011-12												
Model	**A-IVb**			**A-IIIb**			**A-IIb**			**A-Ib**		
	Mean	S.E.	95%CI	Mean	S.E.	95%CI	Mean	S.E.	95%CI	Mean	S.E.	95%CI
Fixed part												
Intercept	-0.088	0.055	(-0.194,0.025)	-0.100	0.035	(-0.169,-0.033)	-0.085	0.052	(-0.185,0.015)	-0.105	0.023	(-0.150,-0.060)
Random part												
Level: network	0.004	0.004	(0.000,0.014)	0.034	0.014	(0.011,0.067)						
Level: district	0.079	0.026	(0.041,0.143)				0.079	0.026	(0.041,0.141)			
Level: school	0.180	0.013	(0.156,0.207)	0.218	0.016	(0.189,0.252)	0.182	0.013	(0.158,0.210)	0.247	0.017	(0.217,0.282)
Level: student	0.763	0.004	(0.755,0.772)	0.763	0.004	(0.755,0.772)	0.763	0.004	(0.755,0.771)	0.763	0.004	(0.755,0.771)
DIC:	175124.9			175126.3			175124.9			175126.6		
pD:	456.2			459.7			456.2			460.7		
Units: network	57			57								
Units: district	32						32					
Units: school	477			477			477			477		
Units: student	68023			68023			68023			68023		
ρ (network ICC)	0.004			0.033								
ρ (district ICC)	0.077						0.077					
ρ (school ICC)	0.175			0.215			0.178			0.245		
ρ (student ICC)	0.743			0.752			0.745			0.755		

5.5.3 School effects measured by the ICC

School ICCs in Models A-Ia (ρ=0.210) and A-Ib (ρ=0.245) are of particular interest, as these models are extended and used as a basis for comparison for subsequent analyses. Here, 21.0% of the total variation in literacy achievement lies between schools, and 79.0% between students within schools; in math, the corresponding proportions are 24.5% and 75.5%, respectively. At this stage, there is already an apparent difference between overall school effects in literacy and those in mathematics. This aligns with the findings of previous school effectiveness studies (e.g. Lenkeit, 2012; Scheerens & Bosker, 1997), which further suggest that this difference between subjects in the magnitude of school effects becomes more pronounced after controlling for the effects of student background and prior attainment.

Gross (unadjusted) school effects from Models A-Ia and A-Ib appear to be quite consistent across subjects. The weighted correlation between these models is strong and positive ($r=0.872$, $p<0.01$), implying that, for the most part, middle schools that are high-performing in mathematics are also high-performing in literacy, and vice versa. Nonetheless, the fact that this correlation is not closer to 1 suggests that some schools may be differentially effective in literacy and mathematics.

5.6 ACCOUNTING FOR STUDENT-LEVEL PREDICTORS: VALUE ADDED MODELS

The results presented in the above sections provide estimates of school performance that do not take school intake into account. In reality, however, students come to a particular school with pre-existing characteristics and prior ability levels; controlling for these student-level predictors disentangles their effects from the effects of the school, providing a measure of the "value added" by a particular school (Timmermans & Thomas, 2014).

5.6.1 Prior achievement

In the literacy model, controlling for prior achievement in literacy led to substantial reduction in the DIC compared to the corresponding empty model (ΔDIC= -11358.3), and the same was true after controlling for prior mathematics achievement in the mathematics model (ΔDIC= -17561.4). These DIC differences far exceed the value of 7 taken to indicate improved model fit. Coefficients for prior literacy achievement in Model B-Ia and prior

mathematics achievement in Model B-Ib were significant, demonstrating that prior achievement in a particular subject, somewhat unsurprisingly, predicted achievement in the same subject in 8th grade. When Models B-Ia and B-Ib are compared to the corresponding empty models, these predictors explained 45.5% and 15.4% of the variation at the school and student levels in the literacy model, and 47.0% and 22.7% of the variation at the school and student levels in mathematics. Tables 11 and 12 show comparisons (including ICCs, total variance, and variance explained) across all stages of model fitting, including the contextual value added models discussed in Sections 5.6.2 and 5.7.

Table 11 Comparison across models, Literacy 2011-12 (ICCs, total variance, and variance explained)

Literacy 2011-12		A-Ia	B-Ia	B-IIa	B-IIIa	C-Ia	D-Ia
Variance	School	0.213	0.116	0.082	0.083	0.046	0.023
	Student	0.800	0.677	0.623	0.616	0.525	0.525
	Total	1.013	0.793	0.705	0.699	0.571	0.548
Variance explained	School		45.5%	61.5%	61.0%	78.4%	89.2%
	Student		15.4%	22.1%	23.0%	34.4%	34.4%
	Total		21.7%	30.4%	31.0%	43.6%	45.9%
ρ (school ICC)		0.210	0.147	0.116	0.119	0.081	0.043

Table 12 Comparison across models, Literacy 2011-12 (ICCs, total variance, and variance explained)

Mathematics 2011-12		A-Ib	B-Ib	B-IIb	B-IIIb	C-Ib	D-Ib
Variance	School	0.247	0.131	0.119	0.119	0.089	0.070
	Student	0.763	0.590	0.575	0.574	0.521	0.521
	Total	1.010	0.721	0.694	0.693	0.610	0.591
Variance explained	School		47.0%	51.8%	51.8%	64.0%	71.7%
	Student		22.7%	24.6%	24.8%	31.7%	31.7%
	Total		28.6%	31.3%	31.4%	39.6%	41.5%
ρ (school ICC)		0.245	0.182	0.172	0.171	0.145	0.119

Additionally, including both subjects as predictors provided a significantly improved model fit over controlling for prior achievement only in the subject of interest (i.e. including prior literacy in the literacy model, or prior mathematics in the mathematics model). Compared to models B-Ia and B-Ib, controlling for prior achievement in the opposite subject further reduced the DIC value by ΔDIC= -5729.0 in the literacy model,

and by ΔDIC= -1759.6 in the mathematics model. These predictors explained an additional 16.0% and 6.7% of the original variance at the school and student levels in literacy, and 4.9% and 2.0% of the original variance at the school and student levels in mathematics, respectively. All of the relevant coefficients were positive and significant, and somewhat unsurprisingly, the effects of prior achievement in the subject of interest were not always larger than those of the other subject in each model. Specifically, in the literacy model (B-IIa) it appeared that prior mathematics score was a stronger predictor of 8th grade literacy achievement than was prior literacy score. In interpreting these results, the reader is reminded that both prior and outcome assessment scores were standardised. Thus, for example, in Model B-IIa, an increase of one standard deviation (1SD) in the standardised 5th grade literacy score is associated with an average increase of 0.267SD in the standardised 8th grade literacy score, while a 1SD increase in the standardised 5th grade math score is associated with an average increase of 0.280SD in the standardised 8th grade literacy score. This seemingly odd result might have reflected the inappropriateness of solely modelling linear relationships, and in fact the addition of a quadratic term for same-subject prior achievement did lead to a less counter-intuitive result.

Tables 13 and 14 display summaries of results from models with coefficients for prior attainment for all three specifications described above.

Table 13 Value added (prior achievement) model results: Literacy, 2011-12

Literacy 2011-12									
Model	**B-Ia**			**B-IIa**			**B-IIIa**		
	Mean	S.E.	95%CI	Mean	S.E.	95%CI	Mean	S.E.	95%CI
Fixed part									
Intercept	-0.041	0.017	(-0.074,-0.009)	-0.021	0.014	(-0.049,0.006)	0.013	0.014	(-0.015,0.040)
Prior literacy	0.379	0.004	(0.372,0.385)	0.267	0.004	(0.260,0.274)	0.319	0.005	(0.311,0.327)
(Prior literacy)^2							-0.035	0.001	(-0.038,-0.032)
Prior math				0.280	0.005	(0.273,0.287)	0.263	0.005	(0.256,0.270)
Random Part									
Level: school	0.116	0.008	(0.102,0.133)	0.082	0.006	(0.071,0.094)	0.083	0.006	(0.072,0.095)
Level: student	0.677	0.005	(0.670,0.684)	0.623	0.004	(0.616,0.629)	0.616	0.004	(0.610,0.623)
DIC:	166971.3			161242.3			160559.4		
pD:	447.6			440.2			442.0		
Units: school	477			477			477		
Units: student	68023			68023			68023		
ρ (school ICC)	0.147			0.116			0.119		
ρ (student ICC)	0.853			0.884			0.881		

Table 14 Value added (prior achievement) model results: Mathematics, 2011-12

Mathematics 2011-12									
Model	**B-Ib**			**B-IIb**			**B-IIIb**		
	Mean	S.E.	95%CI	Mean	S.E.	95%CI	Mean	S.E.	95%CI
Fixed part									
Intercept	-0.062	0.017	(-0.096,-0.028)	-0.059	0.016	(-0.092,-0.028)	-0.065	0.016	(-0.098,-0.034)
Prior math	0.454	0.004	(0.448,0.461)	0.395	0.004	(0.388,0.402)	0.397	0.004	(0.390,0.403)
(Prior math)^2							0.007	0.001	(0.004,0.010)
Prior literacy				0.145	0.004	(0.138,0.152)	0.145	0.004	(0.138,0.151)
Random Part									
Level: school	0.131	0.009	(0.114,0.150)	0.119	0.008	(0.104,0.137)	0.119	0.008	(0.104,0.136)
Level: student	0.590	0.003	(0.583,0.596)	0.575	0.003	(0.568,0.581)	0.574	0.003	(0.568,0.581)
DIC:	157565.2			155805.6			155789.3		
pD:	454.2			453.6			454.5		
Units: school	477			477			477		
Units: student	68023			68023			68023		
ρ (school ICC)	0.182			0.172			0.171		
ρ (student ICC)	0.818			0.828			0.829		

Quadratic terms for prior achievement in the outcome subject for each model were tested and found to be significant in both subjects, although positive in mathematics and negative in literacy. This addition included model fit in both subjects, although substantially more so in literacy, as evidenced by a reductions in DIC (ΔDIC =-683.0 in literacy and ΔDIC= -16.2 in mathematics). However, because additional predictors were not yet controlled for at this stage, the quadratic terms may have been capturing as yet unaccounted-for interactions between prior attainment and other variables. Thus, substantive interpretations of these terms are left for later discussion below.

5.6.1.1 School effects: Value added

The main substantive interest in the present study is how the overall and individual school effects are affected by including these prior achievement variables in the models. In literacy, residual school ICC (i.e. the proportion of unexplained variation that lies between schools) is reduced considerably from that in the null model for literacy after controlling for prior literacy achievement (from 0.210 to 0.147); in mathematics, though the ICC is reduced from that in the empty model, the same is true after controlling for prior mathematics achievement (from 0.245 to 0.182). While in literacy the inclusion of the prior mathematics predictor reduced the ICC further (from 0.147 to 0.116), the addition of a prior literacy predictor in the mathematics model made less of a difference in the ICC (which dropped from 0.182 to 0.172). In fact, as foreshadowed in the previous section, the difference between the literacy and mathematics ICC has become more pronounced after controlling for prior achievement.

While there is still a significant strong positive correlation ($r=0.724$, $p<0.01$) between school effects on literacy and mathematics achievement (calculated using the school-level residuals in the relevant models), there is a marked reduction from the equivalent between-subject school residual correlations based on empty models. This can be interpreted as meaning that there is still considerable consistency across subjects in terms of value added school effects, but this consistency is less apparent after adjusting for prior achievement, pointing to possible differential effectiveness across the two subjects in at least some schools.

5.6.2 Student background characteristics

Numerous studies have shown significant differences in educational achievement for different groups of students; this has led to a strong interest in closing the equity gap in

achievement in many school systems. EER often includes controls for intake differences between schools in effectiveness studies, with the intention of informing "fairer" comparisons between schools. This has sometimes been mistaken for implying that differences related to intake and composition are somehow "excused". On the contrary, the goal of controlling for school intake and composition is to ensure that schools' contribution to the progress of students can be separated from the effects of students' peer group and school population characteristics.

Building upon the models from the previous section, additional student-level predictors were included, with results given in Table 15. There are substantial total reductions in DIC as compared to the prior achievement models B-IIIa and B-IIIb above (with ΔDIC = -10867.1 in literacy, and ΔDIC =-6633.4 in mathematics); in initial processing, there were also incremental reductions in DIC noted with the inclusion of each individual variable in each model. Each of the variables' coefficients are significant in both literacy and mathematics models, with the exception of the coefficients for "Multi" and "Native American" ethnicity categories in mathematics. These categories represent only very small numbers of students, and their variability is proportionately much larger than other ethnicity categories as a result, so substantive interpretations with regard to these specific groups may be problematic. Because "Multi" and "Native American" were defined as categories within the original ethnicity variable, however, they were not removed from the model.

The negative coefficient for gender in both models indicates that on average, boys made less progress than girls, more pronouncedly so in literacy (-0.184SD) than in mathematics (-0.019SD). Similarly, the effects of ethnic categories can be interpreted in comparison to the average performance of White students (the reference category). Thus, in standardised 8th grade literacy, Black and Hispanic students achieved 0.201SD and 0.120SD lower than White students on average, while Asian students' average literacy achievement was 0.069SD higher than their White peers. In mathematics, Black and Hispanic student's achieved on average 0.183SD and 0.124SD lower than White students, respectively, and Asian students outperformed white students, achieving on average 0.368SD higher.

Table 15 Value added (prior achievement and student background) model results

Model	Literacy 2011-12 C-Ia			Mathematics 2011-12 C-Ib		
	Mean	S.E.	95%CI	Mean	S.E.	95%CI
Fixed part						
Intercept	0.466	0.015	(0.435,0.495)	0.158	0.018	(0.122,0.193)
Prior literacy	0.232	0.005	(0.224,0.240)	0.103	0.004	(0.097,0.110)
(Prior literacy)^2	-0.013	0.001	(-0.016,-0.011)			
Prior math	0.188	0.005	(0.181,0.195)	0.316	0.004	(0.309,0.323)
(Prior math)^2				0.014	0.001	(0.011,0.017)
Age	-0.052	0.003	(-0.057,-0.046)	-0.063	0.003	(-0.068,-0.057)
FSM eligibility	-0.144	0.009	(-0.161,-0.127)	-0.058	0.009	(-0.076,-0.041)
ELL status	-0.758	0.010	(-0.776,-0.740)	-0.296	0.010	(-0.314,-0.278)
Ethnicity (Black)	-0.201	0.012	(-0.224,-0.179)	-0.183	0.012	(-0.206,-0.161)
Ethnicity (Hispanic)	-0.120	0.011	(-0.141,-0.099)	-0.124	0.011	(-0.145,-0.104)
Ethnicity (Asian)	0.069	0.012	(0.047,0.092)	0.368	0.011	(0.346,0.390)
Ethnicity (Native American)	-0.218	0.045	(-0.295,-0.140)	-0.070	0.043	(-0.147,0.008)
Ethnicity (Multi)	0.165	0.072	(0.038,0.292)	0.071	0.070	(-0.056,0.197)
Gender (Male)	-0.184	0.006	(-0.195,-0.173)	-0.019	0.006	(-0.030,-0.008)
Disability	-0.365	0.010	(-0.382,-0.348)	-0.365	0.009	(-0.381,-0.348)
Random Part						
Level: school	0.046	0.003	(0.040,0.053)	0.089	0.006	(0.077,0.102)
Level: student	0.525	0.004	(0.520,0.531)	0.521	0.003	(0.515,0.527)
DIC:	149692.3			149155.9		
pD:	434.5			459.3		
Units: school	477			477		
Units: student	68023			68023		
ρ (school ICC)	0.081			0.145		
ρ (student ICC)	0.919			0.855		

In both literacy and mathematics (Models C-Ia and C-Ib), the effects of free or reduced lunch eligibility and special educational needs are negative and significant, with larger effects on literacy achievement than mathematics achievement (particularly in the case of ELL status). The latter might reflect the fact that although ELLs are eligible to receive accommodations for state assessments (e.g. one-to-one or small-group testing environment, extended time, test content read aloud) (NYCDOE, 2014), mathematics assessments still require some degree of reading comprehension or aural comprehension

(whether in English or in another language in translation), and so may measure skills other than mathematics.

The negative coefficients for (standardised) age (-0.052 in literacy, -0.063 in mathematics) suggest that older students have lower average achievement in both subjects. As noted based on the descriptive statistics given in Section 5.4, this negative age coefficient is somewhat surprising, as previous literature suggests that older students within a particular cohort tend to have an advantage (Crosser, 1991; McPhillips & Jordan-Black, 2009; Oshima & Domaleski, 2006). However, the negative age effect found here is likely due to the inclusion of students who were retained in earlier grades, as only those retained during the period studied (i.e. between 5th and 8th grades) were excluded from the sample. In fact, the univariate statistics given in the previous chapter (Section 4.5.2) show that while the average age at the time of 8th grade assessment was approximately 14, the maximum age of students in the sample was over 18.[12] With this in mind, the negative age effect is less surprising.

At the school level, after including the above student predictors, 78.4% of the school level variance and 34.4% of the original student level variance were explained in literacy outcomes. For mathematics outcomes, less of the original variation at the school (64.0%) and student (31.7%) levels were explained. Residual school ICCs representing school effects ($\rho = 0.081$ in literacy, $\rho = 0.145$ in mathematics) were slightly reduced as compared to the previous models in both subjects, demonstrating the importance of controlling for student background in order to draw conclusions about the actual "value added" by schools.

As in the previous sections, consistency across subjects was examined in terms of a weighted correlation. This correlation was still strong and positive ($r=0.703$, $p<0.01$), but lower than the corresponding value obtained at the previous stage of analysis. This suggests that school effects are considerably more variable between subjects after controlling for student background, giving further support to the conclusion that schools had different effects in literacy than they did in mathematics with respect to different groups of students.

[12] Students above the expected age range for 8th grade were not summarily excluded, as school effects calculated based on such as sample would be arguably less meaningful for a school system in which grade retention is commonplace.

5.7 ACCOUNTING FOR SCHOOL CONTEXT AND COMPOSITION: CONTEXTUAL VALUE ADDED MODELS

Several compositional variables were available at the school level. However, initial analyses indicated that the vast majority were not significant. These included proportions by ethnicity, total enrolment, and a categorical variable for school type (i.e. combined elementary and middle grades, middle grades only, or combined middle and high school grades). Proportions of students with disabilities and students with ELL status were available, but these were not used due to changes in policy over the time period considered, as proportions at the school level would risk misinterpretation in later comparisons over time. Additionally, variables related to network affiliation were tested and not found to be statistically significant, including time affiliated (1 or 2 years by the time of the 2012 assessments), total network size (number of schools), network mix (proportions of elementary, middle and high schools), and an indicator for Partnership Support Organization involvement (i.e. whether a network was also associated with an external PSO, or strictly internal to the NYCDOE).

The only compositional or contextual variable with significant effects in the literacy model (Table 16) included school mean intake (5^{th} grade) literacy achievement (centred on the citywide average to facilitate a meaningful interpretation of intercepts). In the mathematics model (Table 17), school mean intake (5^{th} grade) mathematics achievement (also centred on the citywide average) was also significant. Model fit was improved in both subjects by including these predictors when compared to the previous models accounting for student background characteristics (ΔDIC =-44.7 in literacy, ΔDIC =-9.1 in mathematics), though this difference was more pronounced in the literacy model.

Table 16 Contextualised value added model results: Literacy, 2011-12

Literacy 2011-12	D-la		
Model	Mean	S.E.	95%CI
Fixed part			
Intercept	0.478	0.013	(0.452,0.504)
Prior literacy	0.229	0.005	(0.222,0.237)
Prior math	0.185	0.004	(0.178,0.192)
(Prior literacy)^2	-0.013	0.001	(-0.016,-0.011)
Age	-0.051	0.003	(-0.057,-0.045)
FSM eligibility	-0.133	0.009	(-0.151,-0.116)
ELL status	-0.753	0.010	(-0.770,-0.734)
Ethnicity (Black)	-0.187	0.012	(-0.210,-0.165)
Ethnicity (Hispanic)	-0.107	0.011	(-0.127,-0.086)
Ethnicity (Asian)	0.071	0.012	(0.049,0.093)
Ethnicity (Native American)	-0.207	0.044	(-0.285,-0.130)
Ethnicity (Multi)	0.163	0.072	(0.035,0.289)
Gender (Male)	-0.184	0.006	(-0.195,-0.173)
Disability	-0.368	0.010	(-0.385,-0.351)
School Mean Prior Lit.	0.011	0.001	(0.010,0.012)
Random part			
Level: school	0.023	0.002	(0.020,0.027)
Level: student	0.525	0.004	(0.520,0.531)
DIC:	149647.6		
pD:	394.2		
Units: school	477		
Units: student	68023		
ρ (school ICC)	0.043		
ρ (student ICC)	0.957		

Table 17 Contextualised value added model results: Mathematics, 2011-12

Mathematics 2011-12	D-Ib		
Model	Mean	S.E.	95%CI
Fixed part			
Intercept	0.191	0.017	(0.157,0.224)
Prior literacy	0.098	0.021	(0.096,0.109)
Prior math	0.303	0.050	(0.308,0.321)
(Prior math)^2	0.029	0.070	(0.011,0.016)
Age	-0.062	0.003	(-0.068,-0.056)
FSM eligibility	-0.054	0.009	(-0.071,-0.037)
ELL status	-0.294	0.010	(-0.311,-0.275)
Ethnicity (Black)	-0.175	0.012	(-0.198,-0.152)
Ethnicity (Hispanic)	-0.117	0.011	(-0.138,-0.097)
Ethnicity (Asian)	0.369	0.011	(0.347,0.391)
Ethnicity (Native American)	-0.064	0.043	(-0.141,0.013)
Ethnicity (Multi)	0.071	0.071	(-0.057,0.197)
Gender (Male)	-0.018	0.006	(-0.030,-0.008)
Disability	-0.366	0.010	(-0.383,-0.314)
School Mean Prior Math	0.009	0.001	(0.007,0.011)
Random part			
Level: school	0.070	0.005	(0.061,0.080)
Level: student	0.521	0.003	(0.516,0.527)
DIC:	149146.8		
pD:	451.9		
Units: school	477		
Units: student	68023		
ρ (school ICC)	0.119		
ρ (student ICC)	0.881		

The coefficients for the school mean intake achievement variables in each model appear small, though significant. In part, this is a reflection of the use of standardised scores, as school mean prior attainment is calculated in terms of these so as to use the same scale as student prior achievement and outcome variables. For the sake of the present analysis, it is worth remarking that the main concern is less about the relative sizes of fixed effects; rather, the intention is to ensure that significant compositional or contextual variables have been controlled for in order to extract schools' contextual value added effects.

After adjusting for school composition and mean intake achievement, total variance explained at the school level is 89.2% in literacy and 71.7% in mathematics, 10.8% and

7.7% more than in the previous models, respectively. No additional variance is explained at the student level by adjusting for school composition and mean intake achievement; this is as expected, as intuitively school-level variables are not likely to explain within-school variation. School ICCs ($\rho = 0.043$ in literacy, $\rho = 0.119$ in mathematics) are reduced after controlling for school composition and mean intake, and the difference between the two subjects is apparent, as the overall school effect in mathematics is nearly three times larger than the overall school effect in literacy.

Diagnostic checks for the CVA literacy and mathematics models show that residuals are acceptably close to normally distributed at both the school and student levels. Normal probability plots demonstrating this fact are given in Appendix 6.

The weighted correlation between subjects was moderate and positive ($r=0.595$, $p<0.01$). Compared to the corresponding results from the previous section, it is clear that while school effects remain moderately consistent across literacy and mathematics subject areas, they are less so after schools' average intake achievement has been accounted for.

Weighted correlations between residuals from the various models are given in Table 18 below; these provide a useful way to assess the extent to which school effects change in magnitude or direction as a result of adjusting for prior achievement, student background, and school composition. Between-subject comparisons are omitted from the table, as these have already been given in the analyses of consistency in each section.

Weighted correlations between school (shrunken) residuals from empty models and prior attainment value added models, for example, assess the extent to which controlling for prior attainment affected the magnitude and direction of school effects. These correlations were strong and positive ($r=0.961$, $p<0.01$ in literacy and $r=0.947$, $p<0.01$) in mathematics, suggesting that while the effects of individual schools may have changed after adjusting for prior achievement, on the whole the distributions of school effects in empty and prior achievement models were still quite similar. In literacy, the correlations between the residuals at consecutive stages of model fitting were all strong and positive, though decreasing from the earliest to latest stages so that the weakest correlation between effects at one stage and effects at the stage immediately following it compared the student background model (C-Ia) to the school intake model (D-Ia) (0.728, $p<0.01$). This was not strictly the case in mathematics, for which the correlation between residuals at Stages 2 (prior achievement model, B-Ib) and 3 (student background model, C-Ib) was slightly

stronger than that between residuals at Stages 1 and 2 (the empty and prior achievement models, respectively). There was a monotone decrease in the strength of correlations over increasing numbers of steps between the various models, which is unsurprising given that model specifications differed by a greater number of predictors over a greater number of stages of model fitting. The low correlation in literacy between school effects in the initial empty model and the final model including school mean intake achievement (0.363, $p<0.01$) underscores the importance of controlling for the latter in calculations of school effects.

Further, the correlations between empty models and school CVA models are considerably lower than those reported in some previous school effectiveness research in other settings. Based on secondary school achievement data from an inner London borough in England, for example, Strand (2016) found much higher average correlations between similarly specified empty and school CVA models (r=0.78 for English and r=-.82 for mathematics). The lower corresponding correlations given in this chapter suggest that in the New York City context, accounting for school intake achievement substantially changes the rank order of school residuals; while not the substantive focus of this thesis, this finding suggests a possible direction for future investigation.

Table 18 Correlations between school effects from empty, value added, and contextual value added models (Stages 1-4)

		Stage 1		Stage 2		Stage 3		Stage 4	
		Literacy	Math	Literacy	Math	Literacy	Math	Literacy	Math
Stage 1	Literacy	1							
	Math		1						
Stage 2	Literacy	0.961**		1					
	Math		0.947**		1				
Stage 3	Literacy	0.891**		0.947**		1			
	Math		0.864**		0.956**		1		
Stage 4	Literacy	0.363**		0.565**		0.728**		1	
	Math		0.541**		0.759**		0.887**		1

'**' indicates significance at the $p<0.01$ level.

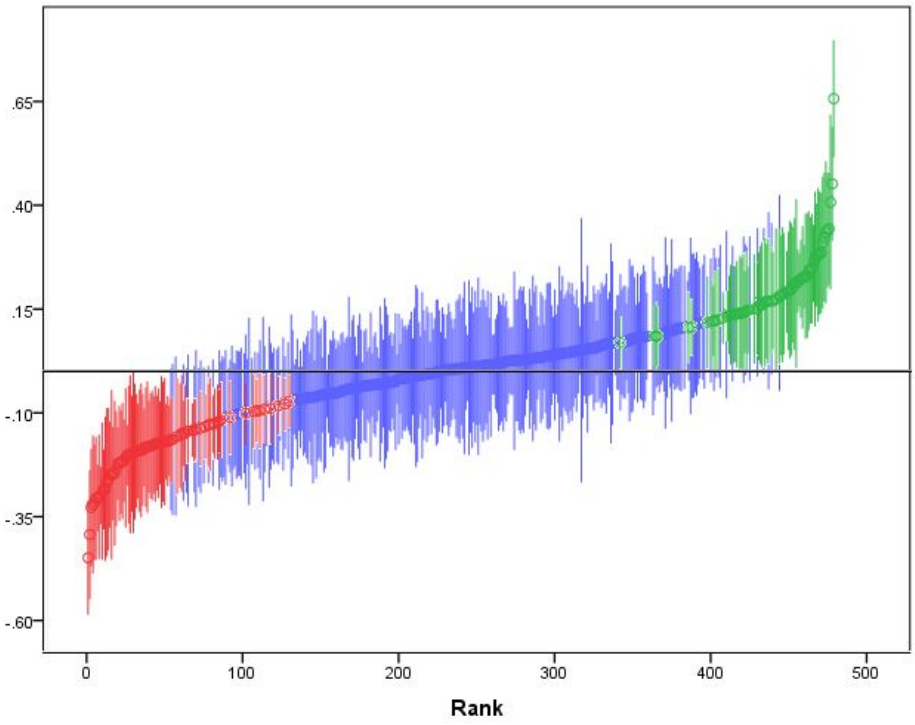
5.8 CLASSIFYING SCHOOL PERFORMANCE

An examination of residuals from the school composition models shows that there is still evidence of significant variation between schools in terms of their effects on literacy and mathematics achievement after controlling for student prior ability, student characteristics and mean school intake achievement. While some of this variation may be due to variables

not controlled for in the models nor included in the available data, these results suggest that schools have significant effects on student achievement over and above the effects of differences in their student composition and intake.

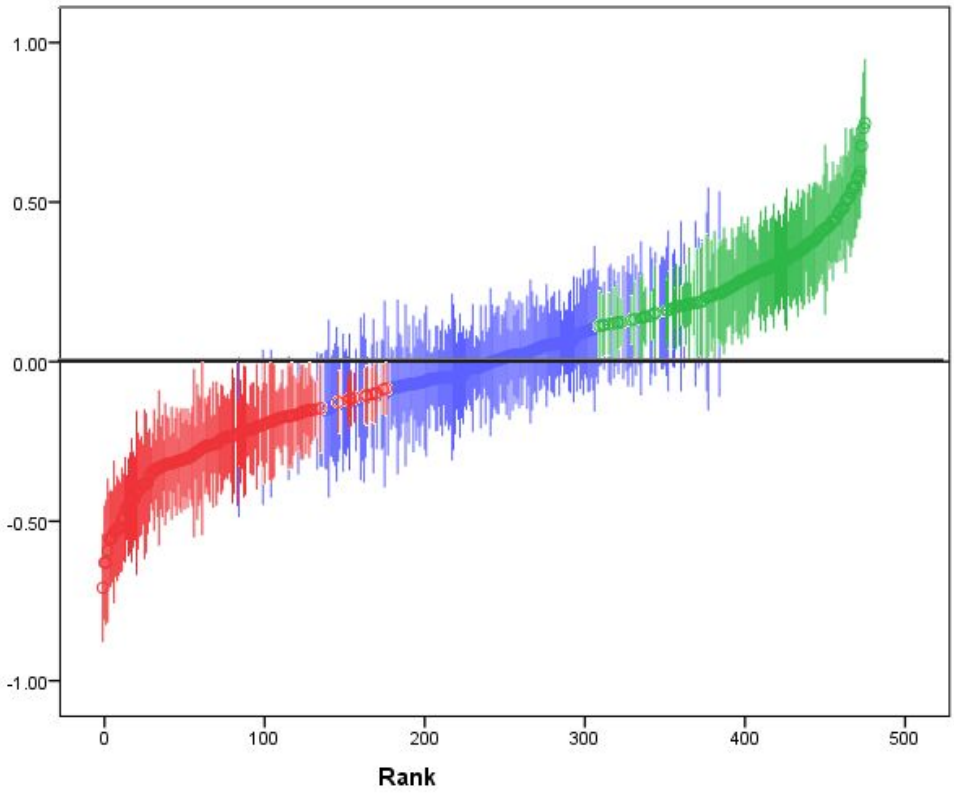
Figures 8 and 9 display caterpillar plots of residuals with their 95% confidence intervals for the literacy and mathematics models, respectively. Those schools for which confidence intervals do not include zero can be interpreted as performing significantly better or worse than expected (according to whether intervals fall above or below the overall zero line).[13] By these criteria, the figures have been highlighted in green for high-performing schools, and in red for low-performing schools, with typically-performing schools displayed in blue.

Figure 8 School (shrunken) residuals and 95% confidence intervals by rank: Literacy, 2011-12



[13] There is an outlier visible on the far right of the literacy plot in Figure 8. To determine whether this individual school was unduly influencing the model results, a separate model was fitted with a dummy variable for this school. Model fit was not significantly improved ($\Delta DIC \approx 1.9$), and no fixed or random effects changed significantly as a result.

Figure 9 School (shrunken) residuals and 95% confidence intervals by rank: Mathematics, 2011-12



Thus, for the purposes of this study, schools are classified as high-performing, typically-performing, or low-performing, depending on these residuals and confidence intervals. Table 19 below shows the number and percentage of schools falling into each category in each subject. It is worth noting that higher proportions of schools are classified as 'high' or 'low' in mathematics than in literacy. This illustrates in discrete categories the meaning of the higher ICC noted above in the final model for mathematics, indicating greater variation between schools in this subject after accounting for student- and school-level variables. The cross-tabulation of schools by subject and performance category in Table 20 additionally demonstrates that only a very small number of schools are high-performing in one subject and low-performing in the other, which resonates with the findings regarding at least moderate consistency from correlation analyses presented above. On the other hand, much larger proportions of schools are classified as low- or high-performing in one subject, and typically-performing in the other. It is worth noting that in such cases, the differences in a school's performance between subjects may actually be quite small (e.g. if its confidence interval for one subject is just below the mean, and the other has just overlapped the mean).

Table 19 Number of schools by performance category in each subject, 2011-12

Performance category	Literacy		Mathematics	
	N	% of Total	N	% of Total
High	83	17.4	141	29.6
Typical	324	67.9	204	42.8
Low	70	14.7	132	27.7

Table 20 Cross-tabulation of schools by performance category and subject 2011-12

		Mathematics		
		Low	Typical	High
Literacy	Low	55 (11.5%)	22 (4.6%)	6 (1.3%)
	Typical	83 (17.4%)	161 (33.8%)	80 (16.8%)
	High	3 (0.6%)	21 (4.4%)	46 (9.6%)

5.9 CONCLUSIONS

Based on the findings presented in this chapter, there is little evidence to suggest an overall effect of networks on student achievement. This does not preclude the possibility of other meaningful differences between networks in relation to alternative outcomes relevant to teaching and learning in schools. It does, however, offer little empirical support for the claim that networks affect student achievement.

On the other hand, the above findings demonstrate significant variation between schools, even after controlling for student and school level predictors. ICCs in each subject are roughly in line with expectations based on the literature, as is the substantially larger overall school effect in mathematics than in literacy (Sammons, 2007; Scheerens & Bosker, 1997). The variability between schools in terms of their CVA effects on achievement, and the differences between subjects with regard to school effects, may have interesting implications with regard to the differentiated support for schools that networks are supposed to provide. Presumably, there is likely to be some correspondence between schools' strengths and weaknesses in promoting student achievement according to the

findings in this chapter, and the support that they seek or receive; this possibility is explored further in light of the qualitative findings in Chapter 6.

Much of the variance between schools was explained by prior achievement in the subject of interest in both literacy and mathematics models, with substantial additional variation explained by controlling for prior achievement in the alternative subject, student background, and mean school intake achievement. The magnitude and direction of some schools' individual effects changed substantially after accounting for student prior achievement and background and school composition. In addition to providing a set of school performance categories for sample recruitment, selection, and description, the CVA results offered a point of departure for some initial understandings of schools' improvement needs. A school with a large positive gross effect but no significant effect based on a CVA model in a particular subject may, for example, have needed improvement in supporting the learning of a specific group or groups of students (e.g. ELLs). This perspective on school improvement needs provides a useful basis for comparison and integration with understandings of school improvement needs and priorities in relation to what support is offered by networks emerging from the qualitative strand of the study in Chapter 6.

Two main findings emerge from the analyses of stability and consistency of school effects presented in this chapter. First, school effects were found to be fairly consistent between literacy and mathematics, based on moderate to strong correlations between residuals in the relevant pairs of models. Second, while both of these statements remained true at each stage of model-fitting, there was a monotone decrease in consistency as additional predictors were included in the models. Thus, schools are very consistent in their gross effects on student achievement, but only moderately consistent in their CVA effects. This has important implications for the present study, taking into consideration the underlying meaning of CVA school effects, which are intended to capture the effects of policies, practices and other dimensions within the control of the school. The fact that CVA school effects are less consistent than gross school effects underscores the importance of taking more than one subject into account in studying school effectiveness.

Although this chapter ends with the models used to inform qualitative sampling, these results also provide a starting point for elaboration and extension in Chapter 7, which

explores the achievement over time and extending the analyses and findings presented in this chapter to include previous and subsequent years (2009-10, 2010-11, and 2012-13).

Finally, the findings presented in this chapter suggest some questions about accountability in context. Networks were not found to have significantly different effects on student achievement in either literacy or mathematics, yet the NYCDOE scores networks on a scale of 1 to 4 that includes an aggregate measure of the performance (on the same state assessments in literacy and mathematics) of their affiliated schools. Further, a report commissioned by the NYCDOE notes that some networks have been closed due to poor performance (Parthenon Group, 2013); it is somewhat unclear whether this refers chiefly to academic performance, but if so, the findings reported in this chapter raise doubts as to the rationale for these closures.

6 NETWORK PARTICIPATION: EXPERIENCES AND PERSPECTIVES

6.1 CHAPTER GUIDE

This chapter is intended to:

- Link findings from interviews to the overall study aims
- Identify the research questioned addressed in the chapter, and explain the approaches taken to address them
- Describe and interpret key themes from interviews with principals, teachers, and network staff
- Discuss alignment of perspectives across stakeholder groups (principals, teachers, and network staff)
- Link with, and elaborate upon, quantitative findings from previous chapters

6.2 CHAPTER INTRODUCTION

Findings in this chapter serve primarily to address the second aim of the study as a whole, that is, to explore the processes and interactions operating within a small selection of networks from the perspectives of various stakeholders (principals, teachers and network staff).

After an account of the approach to analysis, the chapter proceeds with a section presenting findings relevant to the perceived roles of CFNs and the types of support they offer. This is followed by findings relating to networking and collaboration between CFN-affiliated schools, particularly regarding the extent to which CFNs facilitate inter-school collaboration and interaction and the perceptions and values of participants about opportunities for inter-school collaboration. Section 6.6 considers the processes by which practitioners, mainly principals, navigate the CFN structure and the rationales underlying their affiliation choices, as well as the perceived value and impact of CFNs according to participants' accounts. Section 6.7 provides context for participants' perceptions and experiences of CFN participation and support by exploring their views on external pressures, policies and initiatives. Finally, a chapter summary is provided, drawing together key findings from the chapter and linking back to the quantitative results from previous chapters.

6.3 NOTES ON THE ANALYTICAL PROCESS

Findings in this chapter primarily serve to address the following subset of the research questions for this study:

RQ2: *How are the roles and functions of networks understood by various stakeholders (network staff, principals and teachers), and what are the processes and interactions that operate within networks and across the network structure as a whole?*

- ***RQa:*** *How do various stakeholders (network staff, principals, teachers) describe network activities and support offerings, and to what extent do these descriptions align across stakeholder groups?*
- ***RQ2b:*** *To what extent, in what ways, and with whose participation does collaboration and interaction happen within networks?*
- ***RQ2c:*** *What processes and which stakeholders drive network affiliation and shape network support for schools?*
- ***RQ2d:*** *What are the perceptions of various stakeholders about the impact of networks on teaching, learning (e.g. student achievement outcomes), and/or leadership in schools?*

Chapter 4 gives an overview of the approach to analysing qualitative data used in the project as a whole. This section presents the analytical approach specific to the case study component of this research in more depth, specifically with regard to interview data.

As noted in the Methods chapter (Chapter 4), the interview sample consisted of 10 principals, 17 teachers, and 3 network staff members. Because of issues of access also discussed in that chapter, it was not possible to achieve a balanced sample of teachers across participating schools, nor clustered groups of schools within several networks. As a result, findings given in this chapter are interpreted within the framework of the overall CFN structure as the main unit of analysis, with individual schools as subunits of analysis. This is imperfect in terms of supporting generalisability, but does provide exploratory findings to illuminate the processes at work within CFNs, the nature of CFN support, and participants' experiences of participating in these networks.

Transcripts and memos were de-identified with regard to both schools and individuals by assigning each school a letter from A to J, and in cases where multiple teachers in one

school were interviewed, they were assigned numbers in addition to the school letter. The reporting in this chapter uses those identifiers to attribute quotes below.

After interviews were transcribed (including researcher memos taken during interviews as well as interviewee responses), coding was undertaken in several passes using NVivo 10. First, themes were identified based on the questions asked in the interviews, in order to differentiate these themes from emergent themes arising from participants' responses rather than from the framing of the interview schedule. In the sections that follow, the highest level of subheadings represents the broad themes associated with question topics from the interviews. The second round of coding began to identify the emergent themes drawn from participants' responses, and a third pass was aimed at a more fine-grained analysis of sub-themes within the themes and categories defined in previous rounds of coding.

Categories were created to represent school performance (based on results from Chapter 5), job role, and stakeholder experience based on years in current role (up to 4 years was defined as "less experienced", between 5 and 9 years as "middle experienced", and 10 or more years as "more experienced"). Themes and subthemes were subsequently cross-tabulated with these categories to identify any patterns of responses according to these classifications of individuals and schools. There were no clear patterns based on school performance, so that such patterns are not discussed in the sections that follow. However, it was useful to have investigated this possibility, as the lack of patterns according to school performance is indicative in its own right, suggesting that perceptions and experiences of CFN participation were less related to student outcomes in particular schools and more related to other factors discussed in this chapter. On the other hand, some apparent patterns were found according to principals' years of experience, and these are reflected in the relevant sections below.

Another round of coding involved some "pattern matching" (Yin, 2014), drawing on existing theories of networking and school improvement to explore whether any of these seemed to converge with patterns in the interview data. The Effective School lmprovement (ESI) framework that was influential in informing the development of the interview schedule was also a useful tool in defining some of the broad themes as a point of departure for the analysis of interviews; even though defining CFNs as a school improvement device proved to be questionable in context, the ESI dimensions of culture in

terms of shared vision, collegial collaboration, ownership, autonomy, and time (Creemers & Reezigt, 2005), proved highly relevant to participants' accounts of network participation (or lack thereof).

It was apparent from the interview data that principals played a central role in shaping the network structure as the key decision-makers with regard to network affiliation and the most actively engaged school-based participants in CFN-school interactions. Additionally, as a result of the aforementioned complications with regard to access, principals were the only stakeholder group for which there was a balanced sample across participating schools. As such, although this chapter presents findings from principal, teacher and network staff interviews, there is a stronger emphasis on principals' perspectives that is appropriate both to the context and to the nature of the sample.

6.4 DESCRIPTIONS OF NETWORK ROLES AND SUPPORT OFFERINGS

A defining claim about Children First Networks made by the NYCDOE was that they provided broad support for schools. The specific nature of this support and approaches to implementation, however, were not clearly defined in the publicly available documentation. In this section, participants' understandings of the roles of these networks and their accounts of network support are explored based on findings from interviews.

6.4.1 Roles of networks

As noted previously (see Section 1.3 of the Introduction), one of the defining features of CFNs was that they constituted both a network organisational entity with dedicated expert staff, and a network of schools affiliated with this organisation. In their interviews, school-based participants (i.e. teachers and principals) tended to refer to 'networks' as organisational entities rather than as groupings of schools, which demonstrated a key aspect of their understandings of the roles of CFNs. That is, CFNs were most commonly viewed predominantly as support providers rather than facilitators of collaboration, despite the claim of the NYCDOE that one of the underpinning goals of creating the network structure was to help schools collaborate and learn from one another. Network staff were the exception to this; individuals working in formal roles in CFNs did refer more frequently to networks as groups of schools. Collaboration and interconnections within networks is discussed in more depth below in Section 6.5, however, this difference in perception between school-based and network-based staff of what a CFN is at the most

basic level should serve to provide the reader with an initial understanding of some of the tensions and ambiguities surrounding the roles of these networks.

More specifically, three main categories emerged from the thematic analysis to characterise the roles of networks as seen by participants. The first of these conceives of the CFN as a mediator or buffer between schools and the NYCDOE central office and district superintendents. The second depicts CFNs as troubleshooting resources, or safe places to go to ask questions and address immediate issues arising in schools. The third depicts a client-consultant relationship in which CFNs provide a set of support offerings from which schools can pick and choose. Finally, the fourth frames CFNs as collaborating partners in the development and improvement of their affiliated schools and facilitators of relationships (between people, between schools, and with additional external organisation).

All of the principals in the study mentioned the first two conceptions of network roles, although the emphasis placed on these varied according to principal experience. In particular, a much stronger emphasis on networks as buffers and troubleshooting resources was especially apparent in the interviews with the three principals with fewer than five years in their current job roles. These individuals expressed more anxiety about accountability pressures and more of a focus on addressing immediate concerns and questions, so that CFNs were seen mainly as providing answers and solutions on call and keeping principals abreast of changing requirements, deadlines and potential problems. More experienced principals seemed to more commonly view networks in terms of a client-consultant relationship in which it was the job of the network to meet the individual school's needs.

This latter view aligned much more closely with network staff's accounts of their own and their organisations' roles. Network staff spoke largely about the services and supports provided to their affiliated schools and about school improvement planning and collaboration, and mentioned troubleshooting emergent problems and mediating between the central NYCDOE and schools either briefly or not at all.

A few teachers interpreted the role of a CFN mainly as disseminating information. As one teacher explained,

> *"…They [the CFN] would send our principal and assistant principal emails about certain things, and they would forward it to us, and we would look, and if we're interested, we sign up and participate. So they always*

keep us aware of what different…things we can participate in..." (Teacher 2, Science, School J).

It is worth noting that teachers who focused on this aspect of the network's role were concentrated within two of the schools, which may suggest that principals in these schools passed along network communication to teachers either more frequently or in more direct ways than was the case in other schools.

Some of the teachers in the sample seemed to interpret the role of CFNs mainly in terms of instructional support, quite possibly because this was the aspect of CFN support most likely to be visible to teachers. Those who were teaching mathematics and ELA or specialising in Special Education and ELL support tended to give more specific accounts of interactions with their schools' networks in ways that related directly to their subject areas. On the other hand, teachers of other subjects gave accounts that suggested a lack of clarity about the roles and functions of CFNs:

"I mean, every now and then we see people from the network come in for this and that…I've seen a couple here and there, I have no idea who I've seen, or what they – you know, I mean, we see faces that come in and out, and ***I know some of them have been from the network, and some of them have been from other places or whatever. From what I gather, our network is pretty supportive of us…****" (Teacher 2, Art, School J)*

There were formal role boundaries delineating CFN and district responsibilities within the overall NYCDOE school system, but the teacher interviews and some of the principal interviews demonstrated that these boundaries were not clear to everyone working within the system. Further, there were some apparent tensions and conflicts between the networks and districts related to what advice schools were being given and what was being required of them in some instances. As one principal articulated,

*"****The superintendent makes the decision, but the network makes the decisions on what to tell you to do, and sometimes you had a disconnect****. Or when you had to establish goals, or reestablish establish goals, the network would say one thing, and the superintendent wants something else…And…You know, could've been the divide and conquer role that was there, but, you know,* ***when you have a network and you choose the network, and they're coming in to support you, they're going to tell you what you want to hear,*** *and not what you need to hear, in many respects…" (Principal, School A)*

In short, these interview findings demonstrate that understandings of the roles of CFNs within the NYCDOE were somewhat lacking in alignment across stakeholder groups, and often quite different from the stated NYCDOE claims about the purposes of these networks. This lays the groundwork for the sections that follow, as participants' perceptions of CFN roles underpin their accounts of specific network activity, perceived value and impact of CFNs, and collaboration (or lack thereof) between schools in their CFNs.

6.4.2 Types of support

The types of support offered by CFNs, according to the accounts of those interviewed for this study, fell into six main categories. These were:

- Compliance and operations
- Teacher professional development and instructional support
- Leadership development
- Links to external organisations
- Securing funding

Each of these categories is defined and elaborated upon in the subsections below.

6.4.2.1 Compliance and operations

All of the principal interviewees mentioned compliance as a major area in which they received support from their CFNs, and most (8 out of 10) tended to discuss this in the most depth and with the most specificity as compared to other types of network support. The term 'compliance' was used to refer to issues such as meeting deadlines set by the NYCDOE, keeping up with current regulations and policy changes, and preparing for accountability practices such as Quality Review inspections. 'Operations' was generally used to mean considerations involved in the day-to-day running of the school, including physical space, student enrolment, human resources, and similar administrative and management issues. Support for compliance and operations was described as taking several forms, some of which were responsive to individual school circumstances, and others of which were regular and applicable to all schools affiliated with a particular CFN. Regular and planned forms of compliance support included weekly newsletters to principals and monthly principals' meetings which were often centred on a particular compliance issue. Responsive forms of compliance support, on the other hand, included

each principal having a designated 'point person' within their CFN, email alerts based on emergent or potential issues in a particular school, and targeted preparation for Quality Review including a mock review process and dedicated network staff to critique materials prepared for the actual review.

The CFN 'point person' was described as someone who could answer specific questions or connect principals with an appropriate person to answer a specific question. This form of support was seen by principals as responsive and targeted to their needs, as the following quotes indicate:

> *"…The network tries to communicate to me who's responsible for what, and I try to go directly to that person." (Principal, School C)*
>
> *"I can call up and say, I need this, and, you know, or my kids need this, and they're there." (Principal, School E)*

This communication was usually conducted by email or by phone and initiated by principals, but some principals (4 out of 10) also described instances in which network staff had reached out by email or phone to alert them to potential compliance challenges or upcoming deadlines.

Many principals also referred to information dissemination as one of the primary roles of their networks. This was most often accomplished through weekly newsletters to principals about upcoming changes in citywide policy, compliance-related deadlines, and available workshop and professional development offerings through the network, through the central NYCDOE office, and externally.

> *"**The majority of it's compliance**, getting the deadlines met, the secondary thing is the PD support. Now, my current network, they're taking a much greater emphasis on doing the PD, but equally I get calls about making sure I do compliance, like make sure you get this done, make sure you get that done. It's not in an as harassing mode, so it feels better, but it still doesn't change that there's a deadline that we have to meet…" (Principal, School A)*
>
> *Well, they have really helped me….with budgeting. And the management of resources. (Principal, School E)*

Some principals found the amount of information provided by the network more daunting than helpful, or doubted that the network was as "plugged in" to developments within the NYCDOE and New York State as their informal professional networks. This theme was further elaborated in their perspectives on navigating network participation, which is discussed in greater depth in Section 6.6 below.

For the most part, support specific to operations was framed as support for responses to emergent problems in schools. For example, in one school the principal was faced with a situation in which a teacher did not show up to work for a prolonged period of time, and had not had the length of experience nor the training to know how to respond. The CFN provided some support for the principal in dealing with the situation, although she was not satisfied with the support she had received and felt that the CFN could have been more proactive:

> *"So I just assumed that the network would support me, and in the end it didn't, it did not work out that way, and then I…After three years, there's still too much that I do not know, and it's, it is not – doesn't make me feel that great, to be honest." (Principal, School F)*

Teachers and network leaders were less inclined to discuss compliance and operations support, quite possibly because the responsibility and accountability for associated consequences fell mainly to principals. In fact, one of the network staff interviewed specifically stated that while he thought others might claim that CFN support was mainly relevant to compliance and operations, he felt this was inaccurate and that the work of these networks was crucial to building capacity in schools.

6.4.2.2 Teacher professional development and instructional support

Teacher professional development and instructional support was spoken about as one of the most visible aspects of CFN presence in schools for both teachers and principals. This support consisted of a combination of professional development meetings specific to subject areas, and instructional coaching for which subject specialists were sent to schools to work with teachers in their own classrooms.

In many cases, the areas in which schools received area-specific coaching were either literacy, numeracy, or support for groups of students with specialised needs (e.g. special educational needs or language support needs). The frequency of this type of support, as well as the areas for which it was offered, varied between schools. Most principals (7 out of 10) and several teachers (4 out of 17) described coaches' visits as occurring approximately once a month, but in several schools participants mentioned that the frequency of visits had attenuated as schools were perceived as improving. For example, in a school that had been identified as a priority school and subsequently moved out of this category, visits were fewer and further between in the year in which interviews took place:

> *"…**We always got substantive support in our area of need,** which was ELA initially, there was a lot of support coming in, we had – I would say at least coaching twice a month, and it was like, pretty much full day coaching, you know? **That transitioned into, in last year and this year, to more of a special ed focus with special ed coaching, and that's becoming***

> ***more and more sporadic,*** *so that's basically, we get probably a visit from the PSO…every month and a half..." (Principal, School B)*

Principals' accounts of instructional support also indicated that there was some variation between CFNs regarding the extent to which school-based staff were involved in choosing the specific areas in which a school received instructional coaching and teacher professional development. Some principals seemed to describe these choices as pre-determined by network staff, whereas others had explicitly been asked to give input:

> *"Yeah, so in my network we had the option of saying, hey, I'd like instructional support in a minimum of two areas, and so…I received the instructional support for those two areas, and…you know, it was okay." (Principal, School C)*

Of the teachers interviewed, most (10 out of 17) described having little or no contact with network staff nor experience participating in network-led professional development activities beyond the school.

> *"…I haven't had much experience with the network, or, or any contact with them, personally." (Teacher 1, Music, School J)*

Four teachers who had been assigned formal leadership roles in the NYCDOE were an exception; these teachers described attending subject-specific network trainings or activities that they had then been expected to bring back to their schools and share out or "turn-key" to other staff members. In general, many teachers (12 out of 17) saw networks as mainly providing support for principals or administrators, and did not believe that networks had a meaningful effect on their practice or on student outcomes. There was also some uncertainty amongst many of the teachers in the sample (10 out of 17) regarding which professional development opportunities and resources had come from networks rather than from district superintendents or alternative external sources. In view of this lack of clarity regarding the source of specific professional development and instructional support, it is possible that some of the teachers interviewed for the present study may have attended or engaged with such support offerings from their network without knowing that they had done so.

6.4.2.3 Leadership development

Leadership development was not as commonly mentioned as instructional and compliance support by teachers and principals in describing their experiences of CFN participation. In fact, a few (3 out of 10) principals referred specifically to the lack of this type of support or to not being satisfied with what was offered by their CFNs in this regard:

> *"To be totally honest with you, through what I've seen and what I've heard, I mean, the only thing they can provide if they have it is strategic stuff for you with special ed, ELLs, or literacy or math…I don't see how they're cultivating leaders. [laughs] I haven't seen that and I've been in two networks. They may have been trying to, it's not something that…**You know, my own network, I don't know if they have the capacity to cultivate the leadership in there**." (Principal, School A)*

> *"**There is some assumption that if I am going to be, or try to be, or strive to be a school principal, that I had better have it together**, and I'd better be trying to do it…So, yeah, no, a lot less of that [leadership support]" (Principal, School C)*

> *"Well, through the network we definitely – I, so I [sigh] this is interesting. There were a couple years there where my experience with the network was not very helpful to my development, as a school leader, besides sort of the compliance feedback and support." (Principal, School H)*

Teachers did not mention leadership development explicitly, but two teachers did discuss CFN activities in which they had participated that could be interpreted as implicitly fostering teacher leadership. In one case, this involved the teacher attending subject-specific meetings for ELA and then being responsible for building this into professional development for his department within his school. In the other case, the teacher had been sent by his school's principal to attend a meeting about incorporating literacy and mathematics into other content areas, and he had similarly been expected to feed this information back to the staff in his own school.

By contrast, network staff and network leaders interviewed for this research all expressed that they felt their networks offered substantive and tailored leadership support, either via principals' meetings or through individual coaching and mentoring strategies to support new principals or those struggling with specific challenges. The two network staff members interviewed who were not network leaders also mentioned offering opportunities for teacher leadership, and one elaborated on this to describe how teacher leaders were identified jointly between the network and the school in the process of improvement planning so that these teachers could be further supported by participating in network activities and taking on greater roles of informal leadership responsibility in their schools:

> *"…we have different pockets of constituents, like we have our…what we call teacher leaders, that's – it's going to be, ideally one teacher from each school, so a school of 25, a cohort of 25 [correcting self], but we're going to divide that up into five different cohorts, so then the lesson study itself can be a little more intimate…The teacher leaders are identified by the*

> *principal, so we do what we do – leave, we do leave the principal that choice, but we cite criterias of, this has to be a person, that, you know, can take this step, and…conduct at your own school, so there's got to be some leadership role…**And there are folks that we've identified outside of this teacher leadership group that were more…hand-picked in some ways, for lack of word choice, that we invited…They're going to be the ones that will be leading the teacher leaders,** the small cohorts of lesson study. So we work with these cohort of five first, and then they will help us in the network-wide teacher-leader series. **But they're all teachers from our schools, so there are other ways that they can collaborate and share.** The same model kind of follows for our ESL offering, our…special education [pause] support." (Network staff member, affiliated with School F)*

This mismatch between the views of school-based and network-based staff regarding CFN leaedership support offerings seems to indicate either a lack of clarity or information about what support is available within a given network, or a difference between what supports are being described and what is actually being offered in practice. Without direct observation of network activities, no claim can be made in this discussion of findings regarding the reason for the lack of agreement between stakeholder groups, however, it is reasonable to conclude that this is evidence of an issue either in the network structure or in its implementation.

6.4.2.4 Links to external organisations

In addition to the above types of support, which seemed based on participants' accounts to be the core support offerings of CFNs, several of those interviewed mentioned ways in which CFNs had fostered or facilitated partnerships between schools and other external organisations. In some cases these were to support schools in areas of specific need identified by principals, either in a particular subject area or with regard to a particular group of students (e.g. with identified ELL or Special Education learning needs). So, for example, in School B the principal explained that his network leader had connected him with an organisation offering a literacy programme that the school subsequently adopted and found useful for supporting students' reading skills and improving their systems for tracking individual progress. In School D, the principal said that the network had contracted with outside organisations providing specialist support for behaviour management and special education support so that affiliated schools could work with consultants from those organisations to access these resources.

In some cases, these partnerships offered support in areas beyond academics. For instance, in one school staff had identified an ongoing and prevalent issue of students' socio-

emotional challenges, and so had built a relationship with an organisation that provided regular counselling for groups of students, with the initial introduction between the school and the counselling support organisation having been made by CFN personnel. In another case, a network leader described his efforts to connect schools with organisations that provided dance lessons, because he felt that having the chance to engage with the performing arts was important to students' well-rounded development.

6.4.2.5 Securing funding

While not mentioned as ubiquitously as the types of support identified as central to CFN functions, an additional type of support referred to by school-based staff in three of the sample schools was help accessing additional funding. In particular, participants recounted how their affiliated networks had worked with schools to write grant applications. In one instance, this related to a startup grant to help found the school:

> *"…We got [national organisation, name removed] funding, both for the PSO and for us, you know…So, at some level, especially the year or first two years, part of it was that we were getting about, I mean the first two years we got about 500,000 from them.* ***So there was, you know, significant funding that we would get through the Gates foundation through the PSO****." (Principal, School B)*

In another school, the principal was, at the time of the interview, working on a grant application for a special program to support overage eighth-graders to earn high school credits and access individual and family counselling to help them transition successfully out of middle school.

> *"This is what has to be done, this is what has to be worked on. This is the application. So the, the parents would be involved. And that's one of the things that [school number] is very successful with, is bringing parents in and keeping them involved. But that means support – and I'm -- I went off on a little bit of a tangent,* ***but that's something that I spoke to my network about. They've helped me find this grant, and they're helping me write it****." (Principal, School E)*

Of the schools for which grant-writing assistance from their CFNs was mentioned, two were affiliated with PSO-led networks. This is not enough to generalise to the network structure as a whole, but participants' accounts of the support they received suggest at least the possibility that PSO networks may have been more proactive in helping affiliated schools to seek and apply for additional funding than were central NYCDOE CFNs.

6.5 NETWORKING WITHIN NETWORKS: CONNECTING WHOM, AND FOR WHAT PURPOSE?

Another of the explicit claims made about CFNs by the NYCDOE was that they would facilitate collaboration and learning between schools. To address this, one of the goals of the interviews conducted for this dissertation research was to explore the extent to which these networks facilitate interaction and collaboration between affiliated schools, and to clarify the nature of these connections and interactions.

When participants were asked to describe the types of support offered by CFNs, it is worth noting that none spontaneously mentioned facilitating collaboration between schools as such. The findings discussed in this section, therefore, are the result of questions asked specifically about collaboration and interaction. The broad theme of collaboration and interaction between CFN schools, therefore, was not one that emerged organically from participants' accounts, but one that arose from the aims and questions framed at the beginning of the research project and the interview items designed to address those aims and questions. The omission of this topic from participants' responses to questions about types of support and about CFN roles might be interpreted as evidence that facilitating collaboration was not a major feature of the CFN structure, or at least that this was not an aspect of CFNs prioritised by practitioners and network personnel.

Once asked specifically about collaboration between schools, however, participants did discuss several ways in which schools (and their faculties) had opportunities to interact with one another within their CFN groupings. From the interview responses, there appeared to be four main types of interaction and collaboration beyond an individual school:

- Principal to principal
- Principal to network staff
- Teacher to network staff
- Teacher to teacher

These are presented in order according to how frequently (in terms of number of participants) and in how much depth (in terms of time spent in interviews) they were discussed by participants. Thus, connections between principals and between principals and network staff were the main links forming functioning networks of people within the CFN structure, with connections between teachers and between teachers and network staff

either occurring with less density or being regarded as less important by participants based on the weight given to them in interviews.

Based on the descriptions of missions/philosophies in the network directory distributed by the NYCDOE, many CFNs profess to emphasise facilitating collaboration or helping schools to learn from one another as one of their main goals. In most cases this is not made specific, but a few networks' descriptions did mention specific activities involving principals and, to a lesser extent, teachers. Interviews with network staff also suggested that these individuals felt that CFNs were fostering collaboration and interaction. Principal and some teacher accounts did confirm that most of the types of activities mentioned by network staff were made available to them, but the value attributed to collaboration and interaction was not high amongst school-based staff, and these activities were not seen as priorities. This highlights another area of lack of alignment between network-based and school-based personnel.

The four given categories are elaborated in the sections that follow.

6.5.1 Principal to principal

Every principal and all of the network staff members interviewed mentioned monthly principals' meetings as a cornerstone of CFN activity. These meetings varied in focus and purpose, but there was an expectation that principals either attend regularly or, when necessary, send a designee to represent them. Half of the principals described these meetings mainly as a forum for information dissemination, a platform for CFNs to engage with their affiliated principals to discuss recent initiatives and policies to ensure that schools would be compliant with the most recent regulations (for example, at the time of interview one principal had recently attended a meeting with a focus on recent citywide changes to policy concerning Special Education documentation). The other half, sometimes in addition to mentioning information dissemination, described these meetings as opportunities to discuss problems of practice or specific challenges in their schools with their peers; those who emphasised this aspect of monthly meetings also tended to be principals who expressed the most satisfaction with this particular support provision. One principal interviewee underscored this point by describing how he valued engaging with his peers, how the monthly meetings in his particular network had changed, and how he saw himself playing a role in initiating that change:

> *"…And so I had some conversations with the network leader, who was very receptive,* ***we have a good relationship and she was very receptive to these conversations, where I said, these are not helpful meetings, not just for me, but I just don't think they're particularly helpful meetings,*** *you know? And I talked to her about the kinds of professional development that I do, and how it's focused on sort of the adult learning style, and us sharing practices, and us providing support to each other, and talking about the topics that we've really given feedback that we need, not just sort of showing up, and here's what we have for you today…And* ***so, over the past year, they started to…sort of change their approach. And, you know, she made – she definitely acknowledged that our conversations helped that along,*** *and it's become more helpful because when you go and sit with your peers and your colleagues, you really want to talk to them, and you want to hear about their practice, and* ***so now that's more something that happens, and we actually have presented to each other more, and on the topics that we sort of pre-selected, and so on. So that's, that's gotten better****…" (Principal, School H)*

In addition to formally giving principals an opportunity to interact, there was evidence of a relationship between monthly principals' meetings and principals' informal professional networks. There was, however, a sort of "Which came first, the chicken or the egg?" ambiguity to this relationship. Two of the less-experienced principals in the sample related that after being introduced at monthly meetings they had exchanged contact information with more experienced principals, and felt empowered to reach out to those individuals for advice and support. On the other hand, the most and least experienced principals in the sample explained that they had chosen to affiliate or remain with a network that included schools with whom they had a pre-existing relationship, one because of geographical proximity and the other because of shared participation in a pilot initiative (the Empowerment Zone described in the Context chapter, Section 2.3.2), so that the relationships with other principals in their CFNs pre-dated their CFN affiliations.

In addition to principals' meetings, there was some evidence of opportunities for other types of interaction between principals from different schools within a CFN in the form of 'learning walks' or 'lab sites'. These involved principals from affiliated schools receiving invitations to visit a particular school as a group. The purpose of these visits seemed to fall into two broad categories. One of these was for a school to showcase something that it had been identified as doing particularly well (for example, one of the principals shared an invitation he had received to visit a school where students had made greater-than-expected progress in literacy). The other was for principals to visit a school where the principal had identified a specific area in which he or she thought 'critical friends' might be able to

provide advice and feedback to improve practice or solve problems in the host school. One principal described how he had seen differences between networks in the extent to which opportunities for inter-school collaboration and learning were prioritised:

> *"My old network we didn't, we didn't do any of that. We did on our own, but we didn't do the inter-visitations as we should've. In this network, there are meetings in schools, and we do walks, and…The last one,* ***principals presented on best practice, and there were 17 of us out of the 30, and to get – share ideas, and there's principal plus one, which is another person in the school to get feedback on stuff so we can learn from each other, so there's a greater emphasis in this new network on trying to get that done."*** *(Principal, School A)*

The fact that these opportunities were provided, however, did not mean that principals regularly took advantage of them. In fact, the majority of principals (6) did not mention these 'learning walks' or similar activities until specifically asked about them. Even amongst those who did mention these activities without prompting, some principals were reluctant to engage, or selective with respect to collaborating with and learning from particular schools, because they felt that it simply was not of enough benefit to them to justify the time and effort required.

> *"I think for me it was very dependent on who the literacy lab that I was being invited to attend was, who the host was.* ***I didn't particularly think that they had things that I could learn that would be – I don't know, I never went, so I could be completely wrong.****" (Principal, School B)*

> *In part, I know this is going to sound ridiculously immodest, but it's just true, like,* ***when you go to a school that's just nowhere near where your school is, you don't get a lot out of it.*** *You know, like you're look – like, what you, it's – and it's funny, I mean, I've been doing this, I'm going into my tenth year, so I think it's fair to say this at this point. When you walk into a school of somebody who's in their second year as the principal, and they're taking over for a principal who maybe…whatever, maybe they weren't there for very long or whatever…it's just immediately transparent, like, where the pedagogy is at, where the curriculum is at, where the leadership is at, where the organization is at, and I can just sort of see myself six or seven years ago. (Principal, School H)*

This hesitation and doubt about the perceived value of collaborative learning opportunities across schools extended beyond these specific 'learning walk' activities. Based on the interview data, there was evidence of a broader pattern of principals expressing doubt about learning from people beyond their schools. In some cases, this was attributed to differences in experience, while in others it was linked to assertions about the uniqueness of a school's particular composition or needs linked to neighbourhood location.

6.5.2 Principal to network staff

Apart from principal to principal interactions, the most commonly mentioned type of connection between people beyond the boundaries of a particular school was that between principals and network personnel. These connections consisted of:

- A principal reaching out to a "point person" in the CFN for help or information regarding a specific, pressing issue;
- Network staff reaching out via email or phone to remind a principal of upcoming deadlines, requirements, potential compliance issues, or new central NYC DOE mandate; and
- Network staff working with principals to establish support priorities for a given school year.

The first two types of interaction were mentioned at least once in every principal interview, and by two of the three network staff in the sample. Although this was not explicitly framed by participants as a form of networking, it is included here because it demonstrates a key way in which principals were involved in substantive and regular conversation with people outside of their schools in order to address specific needs and concerns. Eight of the ten principals seemed to consider their 'point person' responsive and valued this as a key form of network support that reliably met their needs. As one put it,

> *"So,* ***I believe that the network's purpose is to dissect that information and keep us okay, in line, on track,*** *but that's fine, you still need to have, as any online course, I feel, you need that interaction…So I think the support is in the interaction of having the material reviewed, if there's questions, or other questions that you need answered, it's there." (Principal I)*

Of the remaining two, one had been dissatisfied with her network liason and felt neglected, while the other expressed a preference for relying on informal professional networks for advice and guidance and expressed uncertainty as to whether the network liason was able to consistently answer questions or help with immediate problems.

Regarding outreach from the network by email and phone, most (7) principals suggested that the information received was useful.

> *"And* ***they've really streamline – and I'm lucky that my network sends newsletters, weekly newsletters, of, reminding us what to do,*** *which is, you*

know, with so many compliance things that are going on, it – we, personally being involved in so many other things, it's nice to have that little reminder, you know, next week this is due, are you're working on this, and things of that nature." (Principal, School E)

"I mean, they tell me when the compliance, when, you know, when things aren't happening, when there's new…when there's new directives, and new structures and systems, and we need to be trained, and we need to know how to do them…" (Principal, School H)

The remaining three thought that the amount of information was overwhelming. Further, two felt that the reminders about deadlines and compliance requirements pressurised them and were a source of stress rather than a useful resource.

"A lot of it is… here's your deadline, here's your deadline, here's your deadline." (Principal, School A)

One of the network staff underscored the need for this interaction around support and compliance, even though this individual's main role dealt with instructional support. As she put it,

"I support schools with special ed reforms, transitions for students relevant to their ieps, compliance, assisting schools with policy reforms, writing quality ieps – my work deals with compliance as well as instruction, ***you can't have great instruction without compliance****." (Network staff, CFN affiliated with School D)*

The last type of principal-network interaction was brought up not by principals, but by two network staff members. One, a network leader, described the process by which he worked with individuals to set goals and agree upon priorities for schools:

"So what I've been doing is…today I had two phone calls, I've been calling – ***phone conferences with every principal for them to identify their two greatest – two to three greatest instructional needs****." (Network staff, CFN affiliated with School B)*

The other, an achievement coach, described how improvement planning took place in her network, and the relative roles that schools and network staff played in this process.

"We look at the QR [Quality Review] from the most recent year or past years,
state assessment scores, progress reports, qualitative data we get from being in schools, leadship turnover, school environment scores…some schools have very transient operations, so this can be a challenge. ***Then network staff meet with principals over the summer and have informal conversations about what support is needed, and we don't really get pushback*** *in my experience, principals from schools with the most needs tend to be eager for the support." (Network staff, CFN affiliated with School F)*

The fact that these collaborative improvement planning processes were mentioned by network staff but not by principals could be interpreted as somewhat problematic. One explanation might be that CFNs created opportunities for principals to collaborate with them in planning for improvement and to meet accountability requirements, but not all principals chose to engage with these processes. Another might be that these joint planning efforts were more top-down than implied by network staff accounts, and principals did not feel that there was a collaborative process in place and so did not describe it as such. Regardless of the reason behind it, the fact that network staff gave accounts of a type of collaboration not mentioned by principals highlights a clear discrepancy between stakeholders' perceptions of participation in CFNs as networks.

6.5.3 Teacher to network staff

Most teachers reported little to no interaction with network staff, with the exception of those who had worked with network coaches. Even in these instances, some teachers in the sample expressed a lack of clarity around where support was coming from; they knew that external consultants or coaches had come in to support them in various areas of their classroom practice, but were uncertain as to whether these individuals were from the network, the district, or some other external organisation. In a few instances, teachers had attended network meetings at a principal's request when no administrator was available to attend. Additionaly, two of the teachers in the sample were working as lead teachers in their schools, so that they had attended meetings with colleagues working in other schools in the same subject areas; again, teacher participants were not always sure of whether a particular meeting was run by the network, the local district, or the central NYCDOE.

6.5.4 Teacher to teacher

Most teachers, when asked about whether they had interactions or collaboration beyond their schools, explicitly said that they did not. There were a few (3) exceptions, all teachers who had been given various leadership roles within their schools either in their subject areas or on grade-level teams. These individuals noted that they had attended CFN-led meetings with teachers from other schools. However, even teachers who had attended subject- or topic-specific professional development offered by the CFN, or brought to the school's attention by the CFN, tended to report having stayed within their own department or school groups to facilitate working together on immediate tasks. It was unclear, based

on these teacher accounts, whether facilitating collaboration across schools was a purpose of these meetings; if so, it seemed a purpose that was not fully realised, as teachers did not view these meetings as such.

In one instance, a network instructional specialist mentioned meetings for teacher leaders in which they were encouraged to work together across schools within specific subject areas on tasks that they were subsequently expected to "turn-key" back to either the full staff or specific departments in their schools. When asked about whether her CFN held inter-school collaboration as an intentional goal, she said, *"Yes, yes, definitely…" (Network staff, CFN affiliated with School F).* This certainty, however, seemed not to be reflected in the teacher interviews, as noted above.

Ultimately, it was clear that there was some difference across stakeholder groups in terms of the perceived value and importance of collaboration between schools. The voluntary aspect of participation in specific CFN activities seems to have created a situation in which, based on the interview data, CFNs offering opportunities for collaboration was not sufficient to facilitate sustained and purposeful collaboration and interaction apart from that between principals.

6.6 NAVIGATING AND NEGOTIATING NETWORK SUPPORT

6.6.1 Choosing and changing: Rationales for CFN affiliation choices

Although some NYCDOE documentation suggests that networks should be selected by each school's School Leadership Team, comprised of the principal, parent representatives, teacher representatives and sometimes student representatives, it was clear from principal's responses that for the most part this choice was ultimately made by the principal of each school (choices and the reasons for them, for example, were generally framed as "I" statements). Teachers did not claim to have had a role in choosing networks, and those interviewed for the present study expressed neither an interest nor a feeling of responsibility for participating in this choice. There may have been some bias in the sample of teachers, i.e. those who opted to be interviewed may have been more trusting of their principals and the system in general than those who did not volunteer or refused to participate. Considering the complexities of the system and the fact that networks seem to interact most directly with principals, this is fairly unsurprising. Thus, most of the insight

into reasons for choosing and changing network affiliation draw on interviews with principals rather than other stakeholders.

Principals in the sample for this study gave a variety of reasons for choosing their particular networks, with no apparent patterns based on schools' performance in terms of student outcomes. The main rationales for specific affiliations included:

- Familiar people
- Pre-existing groupings of or relationships between affiliated schools
- Ethos or educational philosophy
- Inherited affiliation from a prior principal

"Familiar people" is used broadly to refer to any rationale drawing on principals' informal professional networks. For example, two of the principals in the sample described following people who they had previously known and trusted in positions of authority.

> *"I went with the person that was my mentor when I first began. And he has a network." (Principal, School F)*

> *"And then when we broke into, when they broke into networks, one of the – the LIST,* ***my lead instructional superintendent…Became a network leader, so I followed her into the network.*** *Then she retired, and another person took over, and I just stayed with the network, because I was just lucky enough that the people who…filled in…you know, replaced the people who retired – we just fit in well. And I think a lot of that may be because of the fact…I know what I want." (Principal, School E)*

In addition to illustrating the "familiar people" rationale for CFN choice, the latter quote also demonstrates the fact that the four types of reasons driving network affiliation given above are not mutually exclusive. In fact, several principals gave more than one reason for their CFN affiliations, or narrated their past experiences of network participation and choice more broadly by referencing more than one rationale.

Not all examples of the "familiar people" rationale were limited to following familiar people who had taken on leadership roles in CFNs. Another way in which principals drew on informal professional networks to inform CFN choice was by relying on word of mouth from other principals, or from former colleagues who had moved into other schools. Not only did several (4) principals mention this in relation to their own affiliation decision-making processes, but one also extended this discussion to explain how former faculty in her school had taken on leadership roles in other schools and subsequently opted in to the same network:

> *"You know, one of my teachers went to become a principal, and he came in our network, one of* ***-- my assistant principal took over a school, and he came back to our network, so there's people that – our schools are already fairly similar…****" (Principal, School G)*

Three principals traced the main reason behind their choices of CFN affiliation to pre-existing groupings of and relationships between schools. In one case, this was based on geographical proximity (i.e. schools within the same district):

> *"First off,* ***the four middle schools around here, we stuck together****. We actually stayed together in the network…So if we had considered moving, we would all have to move together. We, we made that little bit of a pact, why, because we need to be able to exchange information, and we need to support each other." (Principal, School I)*

The two other schools in this category had participated in a particular initiative, the Empowerment Zone, which pre-dated the CFN structure. These principals expressed their sense of a shared understanding between schools that had participated in this initiative, and they had opted to stay in networks along with other schools to which they had been connected through it.

> *"Before my time here,* ***the network was the original Empowerment Zone…So they were a group of schools that came together intentionally in the first place anyway,*** *so I think that has permeated the culture of how the principals work together in this group." (Principal, School C)*

> *"We've had this, we've been in the same network now, I've been in the same one for six years since I've been the principal and the former principal was in the same one, like, I think two years before that. But…****because we were part of the initial Empowerment Zone****." (Principal, School G)*

Another three principals explained their CFN affiliation in terms of shared ethos or educational philosophy between themselves and network leadership. This seemed, based on the accounts given in interviews, to be closely tied to principal autonomy and freedom to do what they thought best for their own schools.

> *"Which is why I joined [the PSO], because* ***they don't come in and tell me what to do, I ask, you know, I tell them what I need, and other organizations don't do that.*** *Other organizations come in a little bit more heavy-handed, and I'm too old and independent at this point to be told, someone telling me what to do, you know? If I've been in successful schools, if I've made schools successful, once I became principal, then I guess I'm doing something right, and I don't want someone telling me to come in and do it this way or that way, since, you know, if it ain't broke, don't fix it." (Principal D)*

Two of the less experienced principals in the sample had opted, in their first year in the principal role, to stay with the networks chosen by their predecessors. One of these principals further explained the reasoning behind this decision, stating that she felt she needed more experience in her new role before being equipped to evaluate whether a change was needed, and there was already enough to which she needed to adapt without also switching networks.

One principal attributed his choice of CFN, which was a PSO-led network, to the fact that it had participated in the founding of the school several years earlier. There was, in this case, some overlap with some of the other reasons given above, as the principal also had a prior professional relationship with the network leader and, implicitly, a sense of shared ethos developed over the years of working together to found and develop the school.

> *"…We've been with the same PSO from the get-go. In fact, we were,* ***they were on my planning team,*** *and we were…I say we, it was a mutual partnership, is the best way of saying it." (Principal, School B)*

Only two of the schools in the sample had changed networks during the tenure of the principal interviewed for this study. In both of these cases, the principals explained their decisions to change networks in terms of mismatched or lacking support. One attributed her decision to a broader sense of a lack of support and neglect, and felt that a formal reprimand she had received from the NYCDOE was at least partially due to her previous CFN's failure to inform and assist her with a challenging operational issue.

> *"And I hold the network totally accountable. Not just the person that they sent, but the person that was over that person, and the person that was over everybody.* ***Were you not reviewing, talking with your new principals? When I'd try to check in with you, after a while, you just really couldn't be bothered.****" (Principal, School F)*

The other principal who had changed networks cited a discrepancy between the perceived capacity of the network and the subject areas in which the school most needed support.

> *"Well, it had to do with the support factor…I always…assumed that…being in a cluster or a network, it was like the way everybody else was. But then as they [colleagues] went into other networks I heard different things going on and…and* ***I realized as I spoke to different principals, it's different.*** *You know, in the cluster that I was in, it was more trying to please a – central, and meeting a deadline, than it was to support the schools, when really what the mission was around what the deadline was for. So…****I mean, I didn't see that we were getting a lot of support, in that area, and I had to look in another direction. Now, we did, and there was a lot of improvements****, in – just the feel, it wasn't like an "I've gotcha" mentality*

> *just moving into this network, and I feel a lot more supported. Now, did ultimately some things change, yes and no, you know we have a lot more people coming in,* ***like this new network has somebody experienced with ESL. It's a huge issue for us. My old network had nobody****." (Principal, School A)*

Network staff provide an alternative perspective on network choice, as CFNs were in a sense competing for schools. One network leader in particular described the tensions involved in this role, as networks are called upon to attract schools and are evaluated by their affiliated schools, but also are responsible for guiding and critiquing principals' and teachers' practices in order to devise the most appropriate allocation of resources and support.

> *"So* ***I had a couple of schools that left my network, and I'm thrilled, because I thought, they're the schools who were not good, and their leaders were worse…And I couldn't work with them, because, you know, they would not accept support even though they were failing.*** *I'm required to provide the support, but in some cases they wouldn't even return emails, they won't return phone calls, so what am I supposed to do with them? And of course, they evaluate me. So let's say in some cases,* ***one of them wanted to do something, which I couldn't deliver, it was just a ridiculous amount of money. And of course he got mad and then wants another network, because he thinks he's going to get that somewhere else. So, you know, it doesn't make sense. So that, yeah, so that's difficult for me, because I tend to be very honest, and I feel that…maybe network leaders are not always as honest as they can be, because they don't want to lose the school.*** *My attitude was, you know, this is what it is, and, you know, I'm not trying to hurt you, I want to help you. And for me it's about the kids, so it bothers me if you don't care about your own kids, and I'm just your network leader, that's a problem. And then I have schools that are great, the principals work with me, and we're able to do wonderful things…"(Network staff, CFN affiliated with School B)*

6.6.2 Negotiating support within networks

As mentioned in Section 6.4 with regard to network roles and types of support, interview responses indicated some variation across networks regarding the extent to which school-based staff had a voice in defining the specific support received by a school. There appeared to be two main approaches to defining CFN support for a given school, one of which was for the principal to initiate requests based on perceived needs, and the other of which was for the network to broadcast a set range of opportunities to all of its affiliated schools and allow principals (and in some cases other members of school faculties) to opt in or out of particular support offerings. A few principals and teachers also seemed to feel

that some types of support were mandatory; this perception went hand-in-hand with a sense of ambiguity surrounding the amount of authority attributed to CFNs.

It was mainly the more experienced principals (4 out of 10) who felt empowered to drive and define their schools' relationships with their affiliated CFNs. This involved opting out of support offerings not deemed useful for the particular school, asking for specific types of support whether or not it had been offered, and offering feedback with the intention of shaping and redefining network activities.

> *"And **I'm not afraid to say, this is what I want, this is what I need**, where – and it may be because I'm, you know, I'm the, a veteran of the system, but it just worked out really well. Because I think, you know, like all principals, you have to know your school, know your students, know your staff, know your parents, and you have to, you know, fight for them. And what you feel is the best, and what may work in my school – and **what, what may work for my school may not work for Mr. [principal name of another person in the room], who's also in the same CFN.** Okay? And I think it's, you know, important that, you know, you know what you need in your school. And what I'm, **what I find is that my CFN is [pause] versatile enough and flexible enough to give what each school may need**." (Principal, School E)*

The majority of teachers expressed that they felt empowered and motivated to seek out opportunities for their own development, and referred to their informal professional networks as sources of information to inform relevant decisions. However, these informal professional networks and sources of professional development were not specific to the CFN affiliated with a particular teacher's school.

> *"I mean, like, I have sought out support in other places, but not necessarily like from within the network." (Teacher 1, ELA, School D)*

As indicated in previous sections, in many cases teachers were also either unaware of the network structure, unsure of which activities involved/came from the network, or sceptical about the relevance of network support to their classroom practice and student needs.

6.6.3 Perceived value and impact of CFNs and the CFN structure

Up to this point, several areas in which there was a lack of alignment between the perceptions and experiences of different stakeholder groups have been identified, along with evidence to suggest that there were discrepancies between the stated purposes of CFNs according to the NYCDOE and their actual implementation. This did not, however, preclude the possibility that CFNs were viewed as useful by many participants. In fact,

most (8) principals spoke about their affiliated CFNs using strongly positive language, using words such as "good", "supportive", and "there for me" to frame their value judgments of network staff and network support.

> *"The network is essentially my go-to for every question I have. So, how to schedule a student, whether to enroll a student, how to enroll a student, an allegation against an adult, an allegation against a student, hiring, firing, evaluating…everything, like,* ***the network is the go-to for everything****." (Principal, School C)*

> *"****I find it very useful****. I find that…A lot of the things that are due, I mean, I've been doing it for ten years, so I know, you know, that this is coming up, this is coming up, so it's nice to have the reminder because sometimes you're just so busy doing the other things, you forget, oh, you know, this is due. And so it's very nice to have that, it really is.* ***And it's nice where you can pull out, you know, pull out your phone and text or call somebody and go…I need this, or I need help.*** *Okay, we'll be right over, or I'll send so and so over tomorrow, okay. And they're very good that way." (Principal, School E)*

Six of the teachers interviewed for this study used similarly positive language in referring to their networks, while the rest remained ambivalent, quite possibly because of the previously noted ambiguity about the specifics of CFN roles, processes and activities amongst teachers.

On the other hand, CFNs were generally not perceived by those in the sample as driving change or improvement in schools. On the contrary, several principals spontaneously volunteered their opinions to contradict this notion, instead expressing a belief that principals and school-based staff were ultimately the drivers of change.

> *"Yeah, no, the network does not so much drive change in the school here. And that's sort of the big catch-22, I think that they're, the network should not be driving change in the school here, because they cannot provide better vision than we can. And I really mean we, like we collectively…****The feedback I get from the network helps to drive change, but they are not sort of the instigating force behind the vision for practice. The instigating force is the relationships that I have, the conversations that I have, the meetings that I go to, the partnerships that I have,*** *the – the conversations that I have internally, and what people see, and, you know, my own sort of interaction with the literature like Danielson and the…and the Common Core standards,* ***and not just mine but the staff's, that's really the driving force****." (Principal, School H)*

This finding is highly relevant to the research questions for this study, as one of the aspects that this research was designed to explore was the extent to which CFN support facilitated school improvement. At least according to school-based participants' accounts, it seemed

that while CFNs were for the most part helpful in supporting the day-to-day work of schools, they had less to do with building capacity for improvement and improving student outcomes.

Network staff, somewhat unsurprisingly, held opposing opinions. The three network personnel interviewed for the present study asserted that CFNs were integrally involved in school improvement processes, and felt that their efforts in this regard were having an effect. One network staff member described what she felt was the result of CFN-led capacity-building in member schools:

> *"As they're planning,* ***they think about the structures and systems first, and then the details of what's going to, like what offering will I provide to my teachers so that it makes the most impact****…So that's – there's definitely changes in that. And there's definitely, and there's coaching around, like, what are you doing on a daily basis, like day-to-day operations." (Network staff, CFN affiliated with School F)*

As previously noted, this evidence of a lack of agreement between network-based and school-based participants' views suggests that either there was a disconnect either between planning and implementation of CFN support, or that there was some lack of understanding, buy-in or engagement on the part of school-based practitioners.

Several principals explicitly stated that, while their network experiences were satisfactory, they felt that the specific structure was irrelevant, and that they as principals would be able to successfully seek out beneficial and essential relationships and resources regardless of the system-wide school support structure.

> *"…I guess just in light of, like, the idea of districts and networks and different things like that****, the more that I think about it, the more I think it's just upon the individual school leader to make those kind of connections, because I don't, I wouldn't necessarily need the network there to have done the things that I did,*** *it's nice that the time is carved out, because of it… You know, I think it's about being really proactive, and, like, what kind of connections do you need to make, you know?...But…I think the overall structure is nice, because there's a lot of freedom in that, and with this empowerment model [referring to networks], that I didn't feel with the district model. The district model, it was sort of like, here's what we're offering and you need to send somebody here today. It was very…it was very, like, everything felt very mandated. Not like if you really needed it." (Principal, School G)*

> *"****Sometimes they [the CFN] give me feedback that helps my vision to evolve. But they don't lead me, per se****. And that really works for me, that works very well for me…because I think that the principal position is*

> *somewhat unique in that you are the leader of your learning community, and so as the leader of you're learning community, you…you're [laughs] – it's almost like a little fiefdom, right? So like, you...it should be democratic in nature, it should be, you know, responsive to the needs of the entire constituency in nature, now the catch-22 is it certainly can't be that, too. I mean, you can have principals that…that whether it be through a lack of capacity or a lack of competency or a lack of vision, or experience…or background or whatever it is, or expertise, really need a more directed vision and a more directed support and guidance.* ***And that sort of gets created or not in the current structure. Sometimes it does, sometimes it doesn't. There's a strong network leader, there's a strong superintendent, there's a strong mentor, somebody's there and they sort of get more directive with that principal. Nobody's been particularly directive with me for a really long time, but I've been very…deliberate about finding the support that I need through different relationships****." (Principal, School H)*

Several principals also indicated that they thought support was offered based on school performance; that is, if a school was doing well according to city and state accountability criteria, the network was likely to do less and interact less with the school. In this framework, being "left alone" by the network was a tacit sign of success.

> *"Mostly what I have them helping me with is…you know, budget, where I'll want to do something that's a little bit different, and they'll tell me how to do it without getting myself fired.* ***Because I – because we are a successful school, I don't need them as much as schools that are struggling****." (Principal D)*

Trust seemed to be a key factor linked to participants' judgments of the value of their affiliated CFNs and the CFN structure in general. Most (7) principals expressed a sense of trust in their current networks when interviewed. As one put it,

> *"It…for me, has felt pretty safe to go to the network because I get to choose the network, so I feel a little like I could…if…when I say safe, I mean – the Department of Ed is very hierarchical, and…yeah.* ***It's easy to go to a network and ask them for help, without…wondering what's the fallout of asking for help****." (Principal, School C)*

Exceptions to this tended to arise when there was a perceived ambiguity or tension associated with the roles of networks. A few (3) principals seemed to view their relationships with CFN staff as more antagonistic, particularly with regard to pressures to comply with regulations and deadlines.

This may have been convoluted by the blurred role boundaries of network staff, many of whom had previously been in supervisory roles within the NYCDOE's accountability

hierarchy (as noted in the section above on reasons for choosing and changing CFN affiliations). That is, the fact that some principals were working with network leaders who had, for example, previously been district superintendents, understandably complicated the relationships and perceptions of "safety" in approaching network staff with conflicts or questions.

One network leader described the difficulties inherent to his role as they related to principals' trust and willingness to engage with the CFN, noting that his ability to support leadership development and ultimately schools' capacity to improve was dependent upon whether principals were open to feedback and constructive criticism. (The quote supporting this was given in Section 6.6.1 above, as it also related to network choice and change.)

For teachers, it seemed to be the case that those who expressed a strong sense of trust in their schools' principals were more inclined to echo this trust in speaking about their affiliated CFNs, even in instances where teachers were uncertain about the specifics of CFN functions and activities.

> *"…As far as, like, our focus is really, again with the [being] academically competitive, is focusing on assessment. And so…and they've been helping us with that". (Teacher, Math, School G)*

Taken together, these findings suggest that while CFNs served a purpose often seen as useful by those practitioners with whom they engaged most directly, there was little evidence to indicate that this purpose included substantive or formal school improvement planning and implementation. Taking into account the views of network staff, it would seem that this was not by design; there was some intention on the part of the CFN organisational entities to engage with schools to plan and facilitate improvement, but at least based on the sample for this study, this intention did not play out in practice.

6.7 PUTTING PERSPECTIVES IN CONTEXT: VIEWS ON EXTERNAL PRESSURES, INITIATIVES, AND POLICIES

As noted in the presentation of context in Chapter 2, the network structure on which this research focused was implemented simultaneously with a number of other significant initiatives and policy changes in the NYCDOE, as well as a highly relevant shift in political control over New York City schools. In discussing their own and their schools' participation in and experience of CFNs, participants tended to refer to a number of these

changes in policy and political context; in many cases, it was clear that certain initiatives and shifts had a considerable effect on individuals' outlooks, motivation and trust in the system overall. Because it was evident during the pilot phase of the study that external pressures, policies, initiatives and political change were closely linked to perceptions about and experiences of CFN participation and support, a question was included at the end of each interview for the main study to give participants the opportunity to discuss these contextual aspects.

Common Core State Standards (CCSS) implementation was a dominant theme, mentioned by almost all of those interviewed regardless of stakeholder group. For principals, key considerations related to Common Core implementation included new assessments and assessment policies, challenges involved in bringing teachers on board and providing them with the necessary training and development to cope with the shift, and lack of clarity around what to expect from new assessment content and accountability. Several principals underscored their relief at the Chancellor's decision not to allow assessments being administered for the first time in 2013-14 to be used to evaluate teachers nor to inform decisions about grade retention for individual students, but there was still some evident trepidation about how results of these assessments would be used in future years. Teachers expressed some of the same concerns about lack of clarity in the rollout of Common Core State Standards and the associated new state assessments. Further, several teachers of subjects other than English Language Arts and Mathematics (for example, an Arts teacher, a Music teacher, and a Science teacher) expressed that they felt that the implementation of CCSS went hand-in-hand with an overemphasis on English and Math at the expense of other subjects both in terms of teaching time (e.g. students being pulled from other subjects for math and literacy interventions and additional instruction) and resources.

> *"…Well, part of my perspective is being a science teacher, and the whole focus being on ELA and math, sometimes I get a little frustrated …Because it's like, all right, I'm trying to do science, like, why do I have to keep sitting through all these things that don't really pertain to me… from what I see here, I feel like this would be, like, citywide, or whatever, that people in my position, teaching a middle school science class aren't really valued, I guess, because they're not – their test scores aren't really going into anything. So…yeah, it's just –* ***it's hard when you're not, when you don't feel like you're as important as the other ones because of the subject that you teach****, and it also makes the kids, because they know, oh, I passed ELA and I passed math, whatever, if I fail science, who cares?" (Teacher 1, Science, School E)*

> *"…It's so, what I've seen is that…[sigh] …I can't blame the administration or anybody for this, but, you know,* ***there's a lot of emphasis on English and Math. So they get a lot of the resources for textbooks.*** *And you only, almost – at least last year they did, this year it seems like they did…textbooks, they have the resources, when we're looking into buying software for students, it's always with Math or, and, and English in mind. And then if it can kind of help out Social Studies and Science, that's a bonus, kind of thing, and then art and music is just, eh [shrugging] type of thing [laughs]…" (Teacher 1, Music, School J)*

On the whole, however, participants took a positive view of the standards themselves, and acknowledged the need for change and rigour. Their objections werer mainly to what was seen as a hasty rollout and lack of clear guidelines to help educators implement the new standards in their classrooms and prepare their students for the new assessments.

> *"The start of a new curriculum without really understanding it…The state is assuming we can just teach it, but how do we learn to teach that?" (Teacher 2, Special Education, School B)*

> *"It seems like a rush job, there's a lack of training." (School Counselor, School B)*

Teacher evaluation systems and procedures were another prevalent theme with regard to external policies and pressures. Similarly to the discussion of Common Core State Standards implementation above, those interviewed generally felt that new rubrics (e.g. the Danielson framework) for teacher evaluation were useful tools for teacher development, but were being misused for evaluation purposes. According to one principal,

> *"I like Danielson, I've used it, we've been using Danielson, this will be our fourth year, so we started using – we use Danielson, and Santa Cruz, teachers had a choice which one they wanted to use.* ***So it's not Danielson, or – the problem is the way the DOE is using it****." (Principal, School D)*

Comments on the new teacher evaluation system and Common Core State Standards rollout were accompanied by many assertions of dissatisfaction with state assessments more generally, in terms both of content and use for accountability and student promotion purposes. Part of this seemed to be a widespread mistrust of the use of numerical data for quantifying teacher effectiveness and student progress in view of perceived complexities inherent to the teaching profession and students' social and emotional contexts.

> *"The focus on testing for accountability, and the new teacher evaluation system with numbers attached that average out numbers…I think teachers overall like how explicit it is, and they like that they get more feedback, that was not a part of the way we were doing things before, but…****Just having the numbers attached, and for…especially at our school where we value people so much – we work with adolescents, and that's what we're***

> ***developing, is the person, not just what they know – we value people so much, to reduce them to numbers is so demoralizing, and it's, it frustrates me, and it – and definitely my teachers, to no end.*** *That's just – to, yeah, to reduce it to the test scores and a number on the evaluation is, it's pretty outrageous. It's pretty unethical, I think, too, actually. I think that's the wrong way of thinking about it. There is no data, no corroborating evidence, to suggest that it actually works, either, you know?" (Principal, School C)*
>
> *"Well, I have great difficulty with the state exams because* ***they're not really testing what they're saying they're testing****." (Principal, School D)*

A few (4) teachers focused on aspects of the school system not related to the broader themes described above, instead emphasising topics of particular relevance to their own professional experiences. In one school, the three teachers interviewed for the study all mentioned discipline procedures in the NYCDOE and expressed that they viewed these as overly punitive. This seemed to reflect a particular focus in the school, as these teachers were not, to the researchers' knowledge, privy to what others had said in their interviews.

A teacher in another school focused his response to the question about external pressures, policies and initiatives on hiring practices, feeling that there was not sufficient diversity and, in particular, that there was not enough emphasis on hiring teachers who were from the specific neighbourhoods and communities served by schools.

These were not, however, topics raised outside of this small selection of interviews.

The proliferation of charter schools was a more commonly mentioned issue. Schools included in the present study were all classified as public schools, most as 'community schools' serving a population of students mainly contained in local neighborhoods within the school's district boundaries, with one theme-based school serving multiple districts with an interest- and performance- based entry requirement. In interviews, several of the teachers and principals mentioned competition with charter schools for resources and for student enrolment as a source of considerable external pressure and tension. One school, for example, was the last remaining public school in its campus building, with all other floors "taken over by charters" (Principal X). A teacher in the school said further that the "best" students were being taken in by charters, leaving their school with a larger proportion of lower-performing students at intake. Concerns about the proliferation of charter schools did not seem to reflect an overall disagreement with the idea of charters, however, some of those interviewed expressed a strong feeling that the need for community schools was being overlooked, and that those in charge of policy decisions

were out of touch with the purpose of community schools and the needs of the student populations they served.

> *"I think though, you know, we need parochial schools, we need private schools, we need public schools, we need charter schools, because you never know what's going to fit a child.* ***What I don't like is that they compare me to charter schools. Charter schools can keep their kids in from 7 to 5…Charter schools also get children whose parents have to opt in, so the minute you have a parent who's invested in the child's education, you're going to get a child who's more invested in his education. That's fine. My school is geared towards helping children who are most in need. And we do it, and we do it well. But don't compare me to a charter school,*** *I can't do what a charter school does. Just hourly. And I'm not getting the same kind of children. Maybe economically, the same kind of children, but the parent involvement is key, all the – you know, you can read research, and more and more research that says, it starts at birth, the more the parents are involved, the education of the parents, the income of the parents, those things are what's going to impact mostly on a child's success at school." (Principal, School D)*

The majority (6) of principals and all three network staff, along with two teachers, mentioned the new Chancellor as representing an important change in the school system. On the whole, feelings about the individual who had recently been appointed to this role were overwhelmingly positive, and this seemed to relate to beliefs that she was an educator by professional background and therefore seen as more knowledgeable and aware of the day-to-day realities and needs of schools and students. These positive views were also given in contrast to a set of commonly-held negative perceptions of the previous political administration (typically referred to as the 'Bloomberg-Klein era', referring to the names of the previous Mayor and Chancellor, respectively).

> *"The last 12 years have been stressful for the system and for teachers. Under Bloomberg and Klein it was really a "gotcha" culture, blaming teachers, oppressive…Since then there have been addresses to that, and positive changes, like disconnecting the reporting of progress to teachers for 2 years while the new test is implemented, and more reasonable teacher evaluations." (Principal, School H)*

> *"It's good to have an educator at the helm. I think for too long we've had business people who want to run the system as a business." (Network staff, CFN affiliated with School B)*

This change in administration was, however, accompanied by a new teachers' union contract about which the views expressed by participants in the sample were less positive. The new contract mandated changes to the processes for parent-teacher conferences and

teacher time allocation that many principals and some teachers saw as impractical or overly restrictive.

> *"Now with this new plan, there's no teacher sup – no teacher-child contact, because one day is professional development and the next day is professional activities and parent contact … Since we have PD built into the day, they don't need to do it after. So we're losing the student-teacher contact. All my interventions – not all my interventions, because we still have the labs during the day, but it's cut in half!" (Principal, School D)*

Views about the new union contract represented just one example of a broader theme in participants' responses, the perception that policy in general was out of sync with the realities needs of students, families, and teachers. This perspective was reflected in a number of teacher, principal, and network staff interviews, and indicated a fairly pervasive cynicism about education policy within the NYCDOE. This often tied back into the previously mentioned concerns about assessment and evaluation:

> *"Yes, we should be measuring performance, and we should be figuring out how our kids are doing, and if people are doing harm. Right? And we should know if people are doing incredible stuff, and I think the tests do tell us that, but to reduce the actual teacher to that number is a whole different thing, that's not using data to inform instruction, that's using data to try to catch people. That's a whole different problem, you know. And, you could even use all that data to set goals and improvement targets, I'm not opposed to any of that, it's just – the reduction of an individual to the number is…too bad, you know?* ***And it doesn't make sense, it's out of touch. Not in this field, not in this work.****" (Principal, School C)*

> *"…I think it was a, I think it was a lost focus, that it was not about kids, for all that they talked about kids first, Children First,* ***you don't put children first by just seeing them as numbers, data…when they don't succeed, you know, then you don't move up. You get this punitive action.*** *And some of that can not be changed by the Chancellor, some of that is just unfortunately the trend that is now, this obsession with testing." (Network staff, CFN affiliated with School B)*

A few (3) of the principals interviewed for the study, notably all of whom had expressed relatively positive value judgments about their affiliated CFNs, highlighted rapid and multiple changes in the school system over approximately the previous ten years as a problem that hindered the potential for meaningful change in schools.

> *"Having been around so long, I can say that change is good but too much change too fast isn't a good thing." (Principal, School J)*

> *"Things change all the time, it takes effort in order to stay current…" (Principal, School D)*

There was some speculation as to whether this had been an intentional effort on the part of the previous political administration, to *"keep people on their toes" (Principal, School H).* Regardless of the motivations behind the rapid progression of policy changes and new initiatives, however, the principals who mentioned this in the context of external pressures also linked their concerns to the CFN structure. This was framed as doubts about whether there were any meaningful differences between this and the previous geographically-based system for support through geographical districts, and expectations that the current structure was also likely to give way to something new before it had a chance to have an effect.

> *"It's, it may be a major miscalculation in terms of your ability to bring out lasting, bring about lasting change. Because* ***you get everybody to essentially adopt an attitude of, well, this change isn't going to stick, because it's not even sticking while we're doing it, so whatever you settled on, no matter how good or bad it was, there's an attitude of, it's just going to change****...You know, whereas if you thought, well I'm going to bring about a structure that's really best for New York City, maybe not do anything for a year or two, and really get a sense of what it is, and then move to the right thing and try to stay there, and maybe tweak it one time, but he came in and changed everything like that [snapping] and then did it every year for like six years, it's crazy." (Principal, School H)*

> *"To me, it's, it is what it's going to be. [laughs] I can't, like, I really don't have much control over those things, or policies, so we just focus on what we can, which is about supporting the schools in any role and capacity.* ***To me, it's like, it doesn't matter what structure, what organization, what's the system, there will be people, that, you know, we'll need, as a system."*** *(Network staff, CFN affiliated with School F)*

6.8 CHAPTER SUMMARY

The interview findings presented in this chapter explain and interpret participants' perceptions of CFN roles and types of support, networking and collaboration within the CFN structure, processes by which practitioners navigate and negotiate network support within this structure, perceived value and impact of CFNs, and external pressures, policies and initiatives contextualising participants' experiences of CFN participation. The lack of distinct patterns based on school performance may be seen as convergent with the quantitative findings, which showed no significant variation at the network level in terms of student assessment outcomes. Meanwhile, the dominant perceptions of network roles and impact suggested that facilitating school improvement was not, at least according to school-based staff, seen as a major function of CFN engagement or activity. This provides

some qualitative confirmation of the quantitative findings, which showed no overall network effect, and offers a possible explanation for this result; if CFNs were not viewed as school improvement mechanisms, it is perhaps quite reasonable that they seem not to have had significant effects in this regard.

Networking and collaboration was also not viewed as a major feature of CFNs' roles by principals and teachers, although there were opportunities for this to take place. The main interactions were those between principals and linking principals to network staff; monthly principals' meetings and interactions with a network 'point person' were generally seen as useful, while opportunities for targeted collaborative learning between schools were often viewed with some scepticism as to their utility and benefit. Overall, networking and collaboration in CFNs was expressed as a priority by network staff but not by principals and teachers, who mostly valued their own informal professional networks over those which might have been facilitated by CFNs.

This represents one aspect of a more prevalent theme of lack of alignment between stakeholder groups in their perceptions of CFN roles that forms a key finding of the present study. This, together with apparent discrepancies between the main claims made about CFNs by the NYCDOE and their implementation in practice, may also help to explain the lack of a significant effect of networks on attainment in the quantitative findings. For CFNs to function as school improvement drivers and facilitators of collaboration between schools, it is possible that a more tightly-focused set of improvement goals within a given network and a more concerted effort to engage practitioners at all levels of the school might have been needed.

That said, CFNs were viewed positively by most of the participants in the sample, and were seen as effective and useful particularly with regard to compliance and operations support and, to a lesser and more variable extent, instructional support. It was mentioned at the beginning of the chapter that a CFN comprised both an organisational entity providing services to schools and a network of affiliated schools; findings strongly indicate that the former characterisation of CFNs was dominant, and that other than a shared affiliation groups of schools affiliated with a given CFN did not function as networks in substantive ways.

7 ACHIEVEMENT IN SCHOOLS AND NETWORKS: INVESTIGATING CHANGE OVER TIME

7.1 CHAPTER GUIDE

The purpose of this chapter is to:

- Answer a subset of the research questions
- Clarify the approach to analysis used to obtain results given in the chapter, including considerations of limitations associated with the data and context
- Connect results to those presented in Chapter 5 on achievement status in 2011-12
- Draw links between the quantitative results given in this chapter and the qualitative findings presented in the previous chapter (Chapter 6).

7.2 CHAPTER INTRODUCTION

Chapter 5 presented results from multilevel analyses of student achievement status using a cross-sectional approach and looking at a 'snapshot' of student and school performance in 2012. This informed the conclusion that there was no overall effect of CFNs, i.e. there was not significant variation at the CFN level in terms of student performance on state assessments over and above the variation attributable to schools and individual students. While this provided an initial indication of the extent to which CFNs might have been associated with achievement, these results were not informative with regard to change over time in schools' performance, and thus could not answer questions relevant to school improvement (or decline).

In order to address whether CFNs were associated with variation in school performance over the first few years in which the network structure was in place in the NYCDOE (beginning after the 2009-10 school year), as well as investigating potential associations with changes in schools' performance during this period, the present chapter builds on the approach and results of Chapter 5 to consider variation in student achievement over time. The approach to analysis used to address this question is discussed in Section 7.3, as are the consequences and limitations to analytical choices of the available data and policy context. Section 7.4 presents results from the empty and final (contextual value added) models from 2009-10, 2011-12 and 2012-13 and compares these to the results from 2011-12 presented in Chapter 5. Section 7.5 presents findings regarding stability of school

effects over the time period considered, and Section 7.6 uses categorisations of schools as high-, typical- and low-performing to examine patterns of schools value added performance over time and whether these appear to be related to network affiliation. The chapter concludes with a summary of findings in Section 7.7.

7.3 NOTES ON THE ANALYTICAL APPROACH

The results presented in this chapter revisit and extend the research questions addressed in Chapter 5 and provides answers to these, using data from the year before and three years during implementation of the CFN network structure in order to explore the issue of stability in ***RQ1c*** and question ***RQ1d.***

- ***RQ1a:*** *How does 8th grade achievement in literacy and mathematics vary between networks, districts, schools, and students, and is there evidence of significant overall effects at these levels?*
- ***RQ1b:*** *To what extent can variation in 8th grade achievement be statistically explained at each level by student background factors, prior attainment in 5th grade, school context and composition, and network or network-affiliation characteristics?*
- ***RQ1c****: How consistent are apparent effects within years across subjects (literacy and mathematics), and how stable are they within subjects across years?*
- ***RQ1d:*** *How do answers to the above questions compare when considering changes (at the school level) in achievement over time?*

Ideally, accounting for change over time would be best done by modelling growth at the student and school level (e.g. Ortega, 2016). However, limitations of the data and of the assessments themselves preclude this approach being used appropriately for the present study. First, the New York State Assessments for grades 3-8 are not vertically equated as noted in the relevant technical manuals (Pearson, 2012a, 2012b). This makes scores on these assessments inappropriate for growth modelling (Singer & Willett, 2003), as direct comparisons of scores over multiple years and at different grade levels would not be meaningful. Second, there was a change in the underlying content and standards as well as the scoring scale and range for the 2013 administration of the assessments, as part of the implementation of the Common Core State Learning Standards in New York. At the same time, CFNs played a crucial role in the Common Core-related policy changes (Wohlstetter

et al., 2015), and excluding this year from consideration in the present study would risk missing important and relevant information.

As an alternative, each year's data was analysed separately using the same approach described in Chapter 5, and then results were compared across years. This involved four main stages. The first of these involved model selection. The second involved assessing stability of effects across years. The third involved categorising schools as high-, typical- and low-performing within each subject for each year. The fourth and final stage involved examining patterns in these categorisations over time and considering whether affiliation with a particular network or networks was related to particular patterns of school improvement, decline, or consistently high or low school performance.

- Model comparisons across years: The results reported in this chapter focus on the empty and final contextualised value added (CVA) models over the four years for which the CFN network structure was fully implemented. For each of these models, intraclass correlations (ICCs) were calculated, representing the proportion of variability at each level (e.g. network, district, school, or student) of a particular model. DIC values were compared to gauge model fit as in Chapter 5, using the rule of thumb that a DIC difference of 7 was sufficient to select one model over another (Spiegelhalter et al., 2002). Predictors were tested for inclusion in each model for each year as before for the 2011-12 models, and results are reported in Section 7.4.2, although the main substantive emphasis, given the research questions, is not on the relative importance of individual predictors but on establishing models that could be compared meaningfully across years and that would produce CVA school effects to be used for subsequent analysis of possible patterns of performance and changes in performance associated with CFN affiliation.
- Stability: Because the modelling approach was undertaken separately for each year within the time period considered (from 2009-10 to 2012-13), it was important to consider the stability of school effects over time. In part, this was particularly useful to support the use of results from 2012 to classify schools as high-, typical-, or low-performing to inform sampling, as these classifications are more meaningful if any school effects identified are found to be relatively stable from one year to the next. Following a procedure identical to that described to investigate consistency across subjects in Chapter 5, correlations

between the school-level residuals obtained from models fitted to data from 2009-10, 2010-11, 2011-12 and 2012-13 were calculated in order to assess the extent to which any school effects were stable over time in literacy and in mathematics. Correlations were weighted as described in Chapter 5 using the inverse of the squared average standard error of each pair of residuals, effectively adjusting for school size. As noted in Section 5.3, the approach of estimating separate models for each subject and subsequently calculating correlations between the two at the school level was considered preferable to estimating multivariate response models. This decision was justified by weighing computation time required against the relevance (to the research questions) of additional information gained by running multivariate response models.

- School performance categories: As in Chapter 5, school-level residuals and their confidence intervals were used to classify each school's performance as high, typical or low in each subject (literacy and mathematics) for each year. Schools with positive residuals and 95% confidence intervals not including zero were classified as high-performing in the relevant subject, schools with negative residuals and 95% confidence intervals not including zero were classified as low-performing in the relevant subject, and all those for which the confidence intervals of residuals included zero were classified as typically-performing. There would be other possible approaches to classify school performance, for example, using ranks of residuals and taking a certain proportion of the highest- and lowest-ranked schools. However, the use of residuals' confidence intervals was considered to be more robust as a measure of whether a particular school's performance was meaningfully (and statistically significantly) above or below average. This strategy is grounded in previous EER literature (e.g. Sammons, Thomas, & Mortimore, 1997).
- Patterns of performance over time: The categories assigned to schools as noted above were used to investigate patterns in performance over one, two and three year time spans. This was first done for all schools to investigate patterns of change over time in the school system overall for middle grades. Possible relationships between CFN affiliation and school performance patterns were then considered for these time spans.

As noted in Chapter 5, MCMC estimation would usually be the preferred approach for calculating parameter estimates for the models presented in this thesis. However, practical considerations including extremely long computation times and limitations of existing software led to the adoption of a more efficient approach similar to that indicated by Browne and Draper (2006). This approach involved beginning with MCMC estimation for empty models in order to make comparisons between non-nested models, followed by the use of likelihood based estimation (iterative generalised least squares (IGLS)) for model exploration and selection with regard to the choice of which level-1 and level-2 predictors to include. The final models were then run using MCMC estimation, with chains of 50,000 and burn-in of 10,000, as investigation of the diagnostics available in MLwiN indicated that these were sufficient for good mixing and convergence.

7.4 MULTILEVEL ANALYSIS RESULTS: 2009-10 TO 2012-13

7.4.1 Partitioning variance

Empty model comparisons for each year in literacy and math were used to establish whether there was statistical evidence of significant variation at the network, district and school levels. As discussed in Chapter 5, comparisons between models with different level and classification structures served two purposes: First, to select whichever model would be most appropriate to use for further analysis and classification of school performance, and second, to address whether networks or geographical districts in fact functioned as meaningful levels within these models. Results reported in this section focus on DIC values and ICCs; tables with full results from all of the empty models can be found in Appendix 5.

The results from the empty models for each year showed that, as for 2011-12, only a very small proportion of the variation in student achievement in literacy and mathematics was at the network and district levels. This was true for three-level models with network or district as the third level, and for four-classification models with schools cross-classified in both networks and districts.

In the 2009-10 model, the year before CFNs were fully implemented throughout the school system, only two level structures were compared for each subject: two-level models with students nested within schools, and three level models with students nested within schools nested within geographical districts. In the three-level literacy model, ICCs indicated that

5.2% of the variation in literacy scores lay between districts, 14.9% between schools, and 79.9% between students. In the three-level mathematics model, 8% of the overall variation in student scores lay between districts, 17.7% between schools, and 74.3% between students. In the two-level literacy model for the same year, 19.5% of the total variation was between schools, and 80.5% between students. In the two-level mathematics model, 24.8% of the total variation was between schools, and 75.2% was between students. Comparing DIC values within each subject between the two- and three-level models indicated that including a district level did not significantly improve model fit (with ΔDIC=-2.2 from the two-level to three-level model in literacy, and ΔDIC=-1.8 for the corresponding comparison in mathematics). DIC values for all empty models are listed in Table 21, below, while ICCs for all empty models for all years are tabulated in Table 22.

For 2010-11 models, which gave results for the first year in which the network structure was fully implemented, four different model structures were compared for each subject: two-level models with students nested within schools, three-level models with students nested within schools nested within districts, three-level models with students nested within schools nested within networks, and cross-classified models with students nested within schools and schools cross-classified in districts and networks at the top level. This was identical to the model structures compared for 2011-12 in Chapter 5. In the cross-classified literacy model, ICCs indicated that 0.5% of the total variation was at the network level, 4.9% at the district level, 14.6% at the school level, and 79.9%[14] at the student level. In the cross-classified mathematics model, only 0.4% of the total variation lay between networks, 6.5% between districts, 15.9% between schools, and 77.3% between students. In the three-level network model for literacy, 2.7% of the total variation lay between networks, 17.3% between schools, and 80.1% between students. In the three-level network model for mathematics, 3.5% of the total variation lay between networks, 18.9% between schools, and 77.6% between students. In the three-level district model for literacy, 5.2% of the total variation lay between districts, 15.0% between schools, and 79.8% between students, while in the corresponding mathematics model 6.7% of the total variation lay between districts, 16.0% between schools, and 77.2% between students. Taken together, it seems that apparent network effects in the three-level models with networks as the top-level units may have actually been reflecting variation between districts, as in both

[14] Note that where total variation appears to differ from 100% by 0.01%, this is merely a consequence of rounding.

subjects the proportion of variation between networks is reduced to less than 1% once districts are included in the level or classification structure. This supports the findings from 2011-12, and provided a first indication that the inclusion of a network level in subsequent models would be inappropriate. In the two-level literacy model for 2010-11, 19.7%% of the total variation was between schools, and 80.3% between students, while in the corresponding mathematics model, 22.0% of the total variation was between schools and 78% was between students. DIC differences for these models in each subject showed that none of the more complex models fit significantly better than the two-level models; the greatest reductions in DIC were ΔDIC=-3.3 in literacy (from the 2-level to cross-classified model) and ΔDIC=-2.4 in mathematics (from the two-level model to both the cross-classified and three-level district models).

For 2012-13, the same four model structures were compared as for 2010-11 and 2011-12. While the ICCs for cross-classified and three-level models in 2010-11 and 2011-12 were quite similar to one another in both subjects, 2012-13 showed slightly higher (though still small) proportions of variation at the network and district levels. In the cross-classified literacy model for this year, 1.7% of the variation was between networks, 7.2% between districts, 14.6% between schools, and 76.5% between students, while in mathematics 1.3% of the total variation was between networks, 10.0% between districts, 17.4% between schools, and 71.3% between students. In the three-level network model for literacy, 4.8% of the total variation lay between networks, 18.4% between schools, and 76.7% between students. In the three-level network model for mathematics, 5.6% of the total variation lay between networks, 22.6% between schools, and 71.8% between students. In the three-level district model for literacy, 7.5% of the total variation lay between networks, 15.8% between districts, and 76.7% between students, while in the three-level district model for mathematics, 10.2% of the total variation was between districts, 18.3% between schools, and 71.4% between students. Finally, for the two-level models for 2012-13, in literacy 22.6% of the total variation lay between schools and 77.4% between students, and in mathematics 27.4% of the total variation lay between schools and 72.6% between students. The slightly higher proportions of variation at the network and district levels in these models may have reflected a consequence of the introduction of Common Core Learning Standards, as qualitative interview findings reported in the previous chapter (Chapter 6) suggested that CFNs as well as districts may have taken on some of the responsibility for supporting the rollout of this curriculum change in schools.

Despite the slightly higher district and network ICCs for three-level and cross-classified models for literacy and mathematics in 2012-13, DIC value comparisons still did not indicate that these more complex models provided a substantially better fit to the data than two-level models using the 7-unit difference rule of thumb. However, using a less stringent rule of thumb for model selection based on DIC differences would suggest that the 2012-13 cross-classified literacy model constituted a borderline improvement over the two-level model (ΔDIC=-5.7), as some authors suggest that a difference of 5 units is sufficient to inform model selection (Browne, 2015). Because the DIC difference could be considered borderline, this cross-classified literacy model for 2012-13 was also extended to include student and school level predictors, with results reported along with the two-level models in Section 7.4.2.

Table 21 DIC and differences across model structures, 2009-10 to 2012-13

		Literacy				Mathematics			
		Cross-classified	3-level (network)	3-level (district)	2-level	Cross-classified	3-level (network)	3-level (district)	2-level
2009-10	DIC			188206.8	188209.0			182301.3	182303.1
	ΔDIC from 2-level model								
2010-11	DIC	182691.0	182693.6	182691.0	182694.3	180707.4	180709.5	180707.4	180709.8
	ΔDIC from 2-level model	-3.3	-0.7	-3.3		-2.4	-0.3	-2.4	
2011-12	DIC	178326.0	178328.4	178326.1	178329.6	175124.9	175126.3	175124.9	175126.6
	ΔDIC from 2-level model	-3.6	-1.2	-3.5		-1.7	-0.3	-1.7	
2012-13	DIC	173523.9	173527.3	173524.9	173529.6	169069.6	169071.5	169070.0	169073.1
	ΔDIC from 2-level model	-5.7	-2.3	-4.7		-3.5	-1.6	-3.1	

Table 22 ICCs for literacy and mathematics models, 2009-10 to 2012-13

		Literacy				Mathematics			
	ICC	Cross-classified	3-level (network)	3-level (district)	2-level	Cross-classified	3-level (network)	3-level (district)	2-level
2009-10	District			0.052				0.080	
	School			0.149	0.195			0.177	0.248
	Student			0.799	0.805			0.743	0.752
2010-11	Network	0.005	0.027			0.004	0.035		
	District	0.049		0.052		0.065		0.067	
	School	0.146	0.173	0.150	0.197	0.159	0.189	0.160	0.220
	Student	0.799	0.801	0.798	0.803	0.773	0.776	0.772	0.780
2011-12	Network	0.008	0.049			0.004	0.033		
	District	0.059		0.060		0.077		0.077	
	School	0.149	0.176	0.157	0.210	0.175	0.215	0.178	0.245
	Student	0.782	0.774	0.784	0.790	0.743	0.752	0.745	0.755
2012-13	Network	0.017	0.048			0.013	0.056		
	District	0.072		0.075		0.010		0.102	
	School	0.146	0.184	0.158	0.226	0.174	0.226	0.183	0.274
	Student	0.765	0.767	0.767	0.774	0.713	0.718	0.714	0.726

These results overall bolster the justification given in Chapter 5 for proceeding in subsequent analyses with a more parsimonious structure of two-level models with students nested within schools. These models were accordingly built upon to include both student and school level predictors according to the same stages and strategy used for the 2011-12 data.

7.4.2 Contextualised value added (CVA) models

As for the analyses presented in Chapter 5, following the comparison of empty models given above, two level models were investigated further for each subject and for each year to establish which predictors were appropriate to include at the student and school levels to produce final CVA models.

Results for the 2009-10 models are given in Table 23, below. All of the student variables tested were found to be associated with literacy and mathematics achievement. These included standardized student age, gender (with Female as the reference category), eligibility for free or reduced school lunch, English Language Learner classification, disability, most categories of ethnicity[15](with White as the reference category), and prior achievement in both subjects. Additionally, a quadratic term for prior achievement in the

[15] Coefficients for 'Multi-Racial' and 'Native American' should be interpreted with caution, as these made up extremely small proportions of the overall population compared to the other ethnic categories.

outcome subject area was found to improve model fit in both literacy and mathematics. At the school level, a contextual measure of school mean intake achievement (calculated based on 5th grade assessment scores and centred on the citywide average) was found to be associated with achievement outcomes in both models. As was the case for the 2011-12 models, other variables relevant to school composition and context were tested but found not to be significant. These included the proportion of students eligible for free or reduced lunch, proportion Black and Hispanic, and grades served (i.e. Middle School (6th to 8th grades), K through 8th, Secondary School (8th through 12th)). Again as in Chapter 5, while proportions of students with disabilities and English Language Learner status were available, these were not used as predictors because changes in inclusion and identification policies over the time period considered meant that these terms could not be relied upon to be comparable in meaning across different years.

The inclusion of student and school level predictors significantly improved model fit, with very large reductions in DIC values of ΔDIC= -23451.3 in the literacy model and ΔDIC= -28886.3 in the mathematics model (model fit based on DIC differences was also checked at each step as each predictor was added to each model). After including these predictors, substantial portions of the variation in the 2009-10 empty models at each level were explained. Specifically, in the literacy model, including student and school level predictors explained 85.8% of the variance at the school level, and 28.1% at the student level. In the math model, 75.5% of the variance at the school level was explained, as was 33.5% at the student level. ICCs calculated based on the remaining variability in these models indicated that 4.6% of the remaining total variation was between schools and 95.4% between students in the literacy model, while in the mathematics model a larger proportion (10.9%) of the total variation lay between schools and 89.1% between students. The finding of greater variation between schools in mathematics rather than literacy results is in accord with most previous EER that indicates school effects tend to be larger for subjects such as mathematics that are primarily learnt in school.

Results of the CVA models for 2010-11 (see Table 24) were quite similar to those reported above for 2009-10. The same set of predictors were found to be significant at the student and school levels, and there was substantial evidence of improvement to model fit based on DIC values (ΔDIC= -28458.4 in literacy and ΔDIC= -22700.4 in mathematics compared to empty models, again with model fit improvement checked with each additional predictor at stages between the empty and full models). Again, much of the

variation at the school level was explained (88.9% in the literacy model, 72.7% in the mathematics model), as was some, although less, at the student level (33.6% in literacy and 27.9% in mathematics). ICCs for these models showed that 3.9% of the remaining total variation in the literacy model was at the school level and 96.1% was at the student level, while in the corresponding model for mathematics 9.6% of the total variation was at the school level and 90.4% was at the student level. As for the 2011-12 models in Chapter 5, variables relevant to network affiliation (e.g. whether a school was affiliated with a PSO or district-run network, density of schools in a particular network, and whether the school changed networks in a given year) were tested for inclusion but were not found to be significant predictors of literacy and mathematics achievement.

Results of the CVA models for 2012-13 (see Table 25) were, for the most part, quite similar in terms of which predictors were associated with student achievement in literacy and mathematics. Interestingly, the main exception to this was that gender was not significant in the mathematics model for 2012-13; this may have been a consequence of some aspect of the new assessments based on the Common Core, although it is not possible to establish a definitive reason for this difference in results based on this analysis or the available data. As was the case for the literacy and mathematics models for previous years, variables relevant to network affiliation were not found to be significant. In the models for 2012-13, DIC differences compared to empty two-level models again showed substantially improved fit (ΔDIC= -28679.2 in literacy and ΔDIC= -29294.2 in mathematics). Adding student and school level variables explained 90% of the variation at the school level and 34.9% of the variation at the student level from the corresponding empty model in literacy, and 80.1% at the school level and 35.6% at the student level in mathematics. ICCs for the CVA models for 2012-13 indicated that 4.3% of the total variation was at the school level and 95.7% at the student level in literacy, and 10.4% of the total variation was at the school level and 89.6% was at the student level in mathematics.

Table 23 Summary of results for CVA models, 2009-10

2009-10	Literacy			Mathematics		
	Mean	S.E.	95%CI	Mean	S.E.	95%CI
Fixed part						
Intercept	0.347	0.015	(0.317,0.376)	0.204	0.017	(0.171,0.236)
Prior lit	0.222	0.005	(0.215,0.229)	0.101	0.004	(0.094,0.107)
Prior math	0.195	0.005	(0.188,0.202)	0.348	0.004	(0.342,0.355)
(Prior literacy)^2	0.020	0.001	(0.017,0.022)			
(Prior math)^2				0.025	0.002	(0.022,0.028)
Age	-0.065	0.003	(-0.071,-0.059)	-0.086	0.003	(-0.092,-0.081)
FSM eligibility	-0.088	0.011	(-0.108,-0.069)	-0.059	0.009	(-0.077,-0.041)
ELL status	-0.568	0.012	(-0.588,-0.548)	-0.173	0.010	(-0.191,-0.155)
Ethnicity (Black)	-0.167	0.013	(-0.191,-0.144)	-0.177	0.012	(-0.199,-0.155)
Ethnicity (Hispanic)	-0.083	0.012	(-0.105,-0.061)	-0.121	0.010	(-0.141,-0.101)
Ethnicity (Asian)	0.077	0.013	(0.053,0.100)	0.360	0.011	(0.338,0.382)
Ethnicity (Nat. Am.)	-0.015	0.055	(-0.119,0.089)	-0.040	0.052	(-0.136,0.057)
Ethnicity (Multi)	-0.067	0.096	(-0.233,0.098)	-0.366	0.090	(-0.518,-0.213)
Gender (Male)	-0.159	0.006	(-0.171,-0.148)	-0.080	0.005	(-0.091,-0.069)
Disability	-0.279	0.010	(-0.298,-0.261)	-0.274	0.010	(-0.291,-0.257)
School Mean Prior	0.008	0.001	(0.007,0.009)	0.007	0.001	(0.006,0.008)
Random part						
Level: school	0.028	0.002	(0.024,0.034)	0.062	0.005	(0.053,0.071)
Level: student	0.595	0.004	(0.588,0.601)	0.506	0.003	(0.501,0.512)
DIC:	164757.62			153416.79		
pD:	386.10			429.59		
Units: school	453			453		
Units: student	70925			70925		
ρ (school ICC)	0.046			0.109		
ρ (student ICC)	0.954			0.891		

Table 24 Summary of results for CVA models, 2010-11

2010-11	Literacy			Mathematics		
	Mean	S.E.	95%CI	Mean	S.E.	95%CI
Fixed part						
Intercept	0.438	0.014	(0.411,0.466)	0.241	0.017	(0.206,0.274)
Prior lit	0.237	0.004	(0.230,0.244)	0.089	0.005	(0.082,0.096)
Prior math	0.187	0.005	(0.180,0.194)	0.302	0.005	(0.295,0.309)
(Prior literacy)^2	0.009	0.001	(0.006,0.011)			
(Prior math)^2				-0.004	0.002	(-0.007,-0.001)
Age	-0.062	0.003	(-0.067,-0.056)	-0.082	0.003	(-0.088,-0.076)
FSM eligibility	-0.102	0.010	(-0.121,-0.084)	-0.056	0.011	(-0.075,-0.036)
ELL status	-0.679	0.011	(-0.697,-0.660)	-0.216	0.011	(-0.234,-0.197)
Ethnicity (Black)	-0.163	0.013	(-0.186,-0.140)	-0.183	0.012	(-0.207,-0.159)
Ethnicity (Hispanic)	-0.107	0.011	(-0.128,-0.086)	-0.123	0.011	(-0.145,-0.102)
Ethnicity (Asian)	0.006	0.013	(-0.017,0.029)	0.303	0.012	(0.280,0.327)
Ethnicity (Nat. Am.)	-0.214	0.046	(-0.298,-0.130)	-0.218	0.046	(-0.305,-0.131)
Ethnicity (Multi)	0.066	0.066	(-0.054,0.185)	0.028	0.065	(-0.097,0.151)
Gender (Male)	-0.193	0.006	(-0.204,-0.182)	-0.033	0.006	(-0.044,-0.021)
Disability	-0.339	0.010	(-0.356,-0.322)	-0.355	0.010	(-0.372,-0.337)
School Mean Prior	0.010	0.001	(0.009,0.011)	0.007	0.001	(0.006,0.008)
Random part						
Level: school	0.022	0.002	(0.019,0.026)	0.061	0.005	(0.053,0.070)
Level: student	0.541	0.003	(0.535,0.546)	0.570	0.003	(0.564,0.577)
DIC:	154235.82			158009.29		
pD:	382.16			435.00		
Units: school	467			467		
Units: student	69216			69216		
ρ (school ICC)	0.039			0.096		
ρ (student ICC)	0.961			0.904		

Table 25 Summary of results for CVA models, 2012-13

2012-13	Literacy			Mathematics		
	Mean	S.E.	95%CI	Mean	S.E.	95%CI
Fixed part						
Intercept	0.433	0.014	(0.408,0.457)	0.261	0.015	(0.231,0.290)
Prior lit	0.251	0.005	(0.242,0.259)	0.099	0.004	(0.093,0.105)
Prior math	0.211	0.005	(0.204,0.218)	0.321	0.004	(0.314,0.327)
(Prior literacy)^2	-0.029	0.001	(-0.032,-0.027)			
(Prior math)^2				0.008	0.001	(0.005,0.011)
Age	-0.060	0.003	(-0.066,-0.054)	-0.061	0.003	(-0.067,-0.056)
FSM eligibility	-0.109	0.007	(-0.123,-0.095)	-0.064	0.007	(-0.077,-0.050)
ELL status	-0.651	0.011	(-0.669,-0.633)	-0.334	0.010	(-0.351,-0.316)
Ethnicity (Black)	-0.209	0.012	(-0.231,-0.186)	-0.236	0.011	(-0.258,-0.213)
Ethnicity (Hispanic)	-0.116	0.011	(-0.136,-0.095)	-0.159	0.011	(-0.179,-0.140)
Ethnicity (Asian)	0.063	0.012	(0.041,0.085)	0.297	0.011	(0.275,0.318)
Ethnicity (Nat. Am.)	-0.185	0.043	(-0.259,-0.111)	-0.093	0.042	(-0.165,-0.022)
Ethnicity (Multi)	-0.038	0.064	(-0.151,0.077)	-0.016	0.063	(-0.125,0.088)
Gender (Male)	-0.096	0.006	(-0.107,-0.085)	-0.003	0.005	(-0.014,0.007)
Disability	-0.415	0.009	(-0.432,-0.399)	-0.428	0.009	(-0.443,-0.413)
School Mean Prior	0.013	0.001	(0.012,0.014)	0.011	0.001	(0.010,0.013)
Random part						
Level: school	0.023	0.005	(0.019,0.027)	0.055	0.004	(0.048,0.063)
Level: student	0.512	0.004	(0.507,0.518)	0.474	0.004	(0.469,0.479)
DIC:	144850.32			139778.96		
pD:	397.02			450.26		
Units: school	483			483		
Units: student	66619			66619		
ρ (school ICC)	0.043			0.104		
ρ (student ICC)	0.957			0.896		

As mentioned in Section 7.4.1, predictors were also added to the 2012-13 cross-classified literacy model to investigate further on the basis that the DIC value was arguably a borderline improvement over a two-level model. Table 26 shows the results of this CVA model. The key finding from this process was that after accounting for student prior achievement, student background, and school mean prior achievement, large proportions of both network and district variation are almost completely explained (83.8% and 97.3%, respectively), leaving extremely low values for variance at these levels. This suggests that what appeared to be network and district effects in the empty cross-classified literacy model may are most likely to have reflected the results of differences between districts and networks in student intake characteristics amongst neighbourhoods and schools (possibly interpreted as selection mechanisms) associated with prior achievement and demographic characteristics. In this model, again as compared to the cross-classified empty model, 87.8% of the variation at the school level and 34.9% at the student level were explained after accounting for student prior achievement and background as well as school mean intake achievement. Resulting ICCs showed 0.5% of the total variation between networks, 0.4% between districts, 3.4% between schools, and 95.7% between students. With a reduction in the DIC value of ΔDIC= -28683.6 as compared to the corresponding empty model, and significant reductions checked with the addition of each predictor, this model was an improvement over the empty model; it was also an improvement over the corresponding two-level CVA model, with a difference in DIC of 10 units. However, with extremely small proportions of the total variation lying between districts and networks, along with less than ideal mixing for the corresponding variance parameters, the two-level models were judged to remain the most appropriate for further analysis.

Table 26 Summary of results for cross-classified CVA literacy model, 2012-13

2012-13	Literacy		
	Mean	S.E.	
Fixed part			
Intercept	0.431	0.017	(0.398,0.464)
Prior ELA	0.251	0.005	(0.242,0.259)
(Prior ELA)^2	-0.029	0.001	(-0.032,-0.027)
Prior Math	0.211	0.005	(0.204,0.218)
Age	-0.060	0.003	(-0.066,-0.054)
FSM eligibility	-0.108	0.007	(-0.122,-0.094)
ELL status	-0.652	0.011	(-0.67,-0.634)
Ethnicity (Black)	-0.206	0.013	(-0.229,-0.184)
Ethnicity (Hispanic)	-0.114	0.011	(-0.135,-0.094)
Ethnicity (Asian)	0.063	0.012	(0.040,0.085)
Ethnicity (Native American)	-0.183	0.042	(-0.258,-0.109)
Ethnicity (Multi)	-0.041	0.064	(-0.155,0.073)
Gender (Male)	-0.096	0.006	(-0.108,-0.085)
Disability	-0.415	0.009	(-0.432,-0.399)
School Mean Prior ELA	0.012	0.001	(0.011,0.014)
Random part			
Level: network	0.003	0.001	(0.001,0.006)
Level: district	0.002	0.001	(0.001,0.005)
Level: school	0.018	0.002	(0.015,0.022)
Level: student	0.512	0.004	(0.507,0.517)
DIC:	144840.3		
pD:	388.6		
Units: network	56		
Units: district	32		
Units: school	483		
Units: student	66619		
ρ (network ICC)	0.005		
ρ (district ICC)	0.004		
ρ (school ICC)	0.034		
ρ (student ICC)	0.957		

Overall, the results given in this section suggest that, with the exception of the uniquely non-significant gender coefficient in the 2012-13 mathematics model, the results were quite similar across the different years when the effects of including student prior achievement, student background and school intake achievement means had been accounted for. It was these CVA models that were used for subsequent classifications of

school performance by year and school performance patterns over time discussed in the sections that follow.

7.5 STABILITY OVER TIME

Tables 27 and 28 below display the correlations between school residuals from different years from 2009-10 to 2012-13 in literacy and mathematics for the empty models. As these results show, correlations between years are strong, positive and significant at the $p<0.01$ level, even for the largest timespan (from 2009-10 to 2012-13). This suggests that schools' performance in terms of state assessment results was very stable across these years, even taking into consideration the change in assessments for 2012-13. The weakest correlation was in mathematics between 2010-11 and 2012-13, still only slightly below 0.9 and significant ($p<0.01$).

For these and all subsequent correlations between pairs of years, only those schools that were open in both years could be included in the analysis. The numbers of schools for each pair of years was as follows:

- 2009-10 and 2010-11, 448 schools
- 2009-10 and 2011-12, 436 schools
- 2009-10 and 2012-13, 434 schools
- 2010-11 and 2011-12, 454 schools
- 2010-11 and 2012-13, 453 schools
- 2011-12 and 2012-13, 475 schools

Table 27 Stability of correlations of residuals at the school level: 2009-10 to 2012-13 Literacy, empty models

Literacy	2009-10	2010-11	2011-12	2012-13
2009-10				
2010-11	.922**			
2011-12	.923**	.941**		
2012-13	.912**	.933**	.947**	

Pearson correlations were used. '**' indicates significance at the $p<0.01$ level (2-tailed).

Table 28 Stability of correlations of residuals at the school level: 2009-10 to 2012-13 Mathematics, empty models

Mathematics				
	2009-10	2010-11	2011-12	2012-13
2009-10				
2010-11	.917**			
2011-12	.900**	.914**		
2012-13	.893**	.907**	.925**	

Pearson correlations were used. '**' indicates significance at the p<0.01 level (2-tailed).

Tables 29 and 30 show the correlations between school residuals from the final CVA models for the same years as above in literacy and mathematics. As compared to the correlations reported above for the empty models in these subjects for the same years, school effects appear much less stable after accounting for student prior achievement, student background characteristics, and school average intake achievement. Nonetheless, most correlations were still moderate and significant (p<0.01), with the exception of the correlation between 2009-10 and 2012-13 literacy school effects, which was significant and positive but relatively weak (r=0.373, p<0.01).

These stability correlations strike something of a contrast with research conducted in England. For example, Strand (2016), found much less stability over one and two year time spans (2004-5, 2005-6, and 2004-6) in empty models for English (r=0.75, r=0.60, and r=0.63, respectively) and slightly less stability in mathematics (r=0.89, r=0.89, and r=0.80, respectively). The same study also found, for the same time spans, considerably lower stability correlations in English (r=0.38, r=0.27, and r=0.23) but mostly similar – and more widely variable – stability in mathematics (r=0.70, r=0.77, and r=0.58, respectively) as compared to the results in Tables 29 and 30 for one- and two-year comparisons. Differences in stability between literacy achievement in New York City and English achievement in an inner London setting might be due to differences in the specific composition (and changes over time in that composition) of the population of students overall, perhaps particularly of the group of students classified as English Language Learners (or, in England, students with English as an additional language), or with the nature of and changes to the educational provision for groups of students with special educational needs (including English language support). The reasons behind these differences between contexts are beyond the scope of this thesis, but might constitute a direction for future enquiry.

Table 29 Stability of correlations of residuals at the school level : 2009-10 to 2012-13 Literacy, CVA models

Literacy				
	2009-10	2010-11	2011-12	2012-13
2009-10				
2010-11	.593**			
2011-12	.509**	.617**		
2012-13	.373**	.498**	.520**	

Pearson correlations were used. '**' indicates significance at the p<0.01 level (2-tailed).

Table 30 Stability of correlations of residuals at the school level: 2009-10 to 2012-13 Mathematics, CVA models

Mathematics				
	2009-10	2010-11	2011-12	2012-13
2009-10				
2010-11	.710**			
2011-12	.611**	.681**		
2012-13	.493**	.588**	.661**	

Pearson correlations were used. '**' indicates significance at the p<0.01 level (2-tailed).

The finding that most of these correlations between years are significant and at least moderate lends support for the use of the 2011-12 results to inform sampling choices for the qualitative strand of the study, in view of the lack of appropriate data that would allow for growth modelling of multiple years' assessment results to be modelled together to reflect achievement growth at the student and school levels.

Consistency of school effects was also revisited using the same strategy as above across all years considered. The results of this analysis are shown in Tables 31 and 32 below, which gives correlations between subjects by year. These also show strong positive and significant (p<0.01) weighted correlations between subjects in the empty models for each year, and moderate but significant (p<0.01) weighted correlations between subjects in the CVA models for each year. This suggests that while unadjusted or 'raw' school effects tend to be very similar across literacy and mathematics, this relationship is less strong after controlling for student background characteristics, student prior achievement, and average school intake achievement. This in turn underscores the importance of considering student attainment and CVA progress in both subjects to answer questions about school performance based on student achievement results.

Table 31 Consistency of correlations between literacy and mathematics residual estimates for schools, empty models 2009-10 to 2012-13

		Literacy			
		2009-10	2010-11	2011-12	2012-13
Mathematics	2009-10	.880**			
	2010-11		.858**		
	2011-12			.872**	
	2012-13				.907**

Pearson correlations were used. '**' indicates significance at the p<0.01 level (2-tailed).

Table 32 Consistency of correlations between literacy and mathematics residual estimates for schools, CVA models, 2009-10 to 2012-13

		Literacy			
		2009-10	2010-11	2011-12	2012-13
Mathematics	2009-10	.529**			
	2010-11		.556**		
	2011-12			.595**	
	2012-13				.571**

Pearson correlations were used. '**' indicates significance at the p<0.01 level (2-tailed).

7.6 PATTERNS OF HIGH, TYPICAL AND LOW SCHOOL PERFORMANCE OVER TIME

Using the CVA two-level models, schools' performance was categorised as high, typical or low in each subject for each year as described in the notes on the analytical approach in Section 7.3. Before examining patterns, overall counts and percentages of schools falling into each category for each year are given in Table 33. These results show that the proportions of schools categorised as high-, typical- and low-performing are fairly stable for each subject over all years from 2009-10 to 2012-13, and that the proportions of schools falling into the high and low categories are relatively balanced within each year for each subject. Additionally, it is apparent that more schools (approximately two-thirds for each year considered) fall into the 'typical' category in literacy than in mathematics (for which less than half of all schools are classified as typically-performing for each year). This difference between subjects is unsurprising, given the larger proportions of variation in the multilevel CVA models presented above as well as the finding that less of the school-level variation was statistically explained in mathematics than in literacy after

accounting for student prior achievement and background and school mean intake achievement.

Table 33 Counts and percentages of schools categorised as high-, typical- and low performing for each year: 2009-10 to 2012-13

		Performance category						
			Low		Typical		High	
	Year	Total schools	Freq.	%	Freq.	%	Freq.	%
Literacy	2009-10	453	79	17.4	304	67.1	70	15.45
	2010-11	467	79	16.9	314	67.2	74	15.85
	2011-12	477	83	17.4	324	67.9	70	14.68
	2012-13	483	84	17.4	321	66.5	78	16.15
Mathematics	2009-10	453	126	27.8	203	44.8	124	27.37
	2010-11	467	128	27.4	220	47.1	119	25.48
	2011-12	477	141	29.6	204	42.8	132	27.67
	2012-13	483	127	26.3	236	48.9	120	24.84

7.6.1 Patterns over 1-year time spans

As an initial investigation of patterns over time, school performance categories were cross-tabulated for each pair of years, with results given in Tables 34 through 36 for literacy and 37 through 39 for mathematics. The total number of schools for each pair of consecutive years is different from the numbers of schools shown in the empty and CVA models above for each year due to school closures and other such events affecting the availability of school data from one year to the next; each cross-tabulation includes only those schools that were open and serving middle grades for both of the relevant years.

Results in literacy show that over each one-year time span, the largest proportion of schools were classified as consistently ‘typical’ (51.6% of schools from 2009-10 to 2010-11, 50.5% of schools from 2010-11 to 2011-12, and 48.6% of schools from 2011-12 to 2012-13). The smallest proportions of schools for each one-year comparison were those classified as ‘low’ in the first year and ‘high’ in the following year (0.9% from 2009-10 to 2010-11, 0.2% from 2010-11 to 2011-12, and none from 2011-12 to 2012-13) and those classified as ‘high’ in the first year and ‘low’ in the subsequent year (0.9% from 2009-10 to 2010-11, 0.4% from 2010-11 to 2011-12, and 0.4% from 2011-12 to 2012-13).

Somewhat larger proportions of schools showed improvement from 'low' to 'typical' (7.8% from 2009-10 to 2010-11, 8.8% from 2010-11 to 2011-12, and 10.3% from 2011-12 to 2012-13) and from 'typical' to 'high'(7.4% from 2009-10 to 2010-11, 6.8% from 2010-11 to 2011-12, and 9.5% from 2011-12 to 2012-13), and similar proportions of schools showed decline from 'high' to 'typical' (6.9% from 2009-10 to 2010-11, 8.4% from 2010-11 to 2011-12, and 7.8% from 2011-12 to 2012-13) and from 'typical' to 'low' (8.3% from 2009-10 to 2010-11, 9.2% from 2010-11 to 2011-12, and 9.7% from 2011-12 to 2012-13) over each one-year period. Proportions of schools classified as consistently 'high' (7.8% from 2009-10 to 2010-11, 7.3% from 2010-11 to 2011-12, and 6.5% from 2011-12 to 2012-13)or consistently 'low' (8.5% from 2009-10 to 2010-11, 8.4% from 2010-11 to 2011-12, and 7.2% from 2011-12 to 2012-13) over each one-year period were relatively balanced and fairly stable across the different single-year comparisons.

Table 34 Cross-tabulation of school performance categories, 2009-10 to 2010-11 (Literacy)

Literacy			2010-11			
			Low	Typical	High	Total
2009-10	Low	Count	38	35	4	77
		% of Total	8.5%	7.8%	.9%	17.2%
	Typical	Count	37	231	33	301
		% of Total	8.3%	51.6%	7.4%	67.2%
	High	Count	4	31	35	70
		% of Total	.9%	6.9%	7.8%	15.6%
	Total	Count	79	297	72	448
		% of Total	17.6%	66.3%	16.1%	100.0%

Table 35 Cross-tabulation of school performance categories, 2010-11 to 2011-12 (Literacy)

Literacy			2011-12			
			Low	Typical	High	Total
2010-11	Low	Count	38	40	1	79
		% of Total	8.4%	8.8%	.2%	17.4%
	Typical	Count	42	230	31	303
		% of Total	9.2%	50.5%	6.8%	66.6%
	High	Count	2	38	33	73
		% of Total	.4%	8.4%	7.3%	16.0%
	Total	Count	82	308	65	455
		% of Total	18.0%	67.7%	14.3%	100.0%

Table 36 Cross-tabulation of school performance categories, 2011-12 to 2012-13 (Literacy)

Literacy			2012-13			
			Low	Typical	High	Total
2011-12	Low	Count	34	49	0	83
		% of Total	7.2%	10.3%	.0%	17.5%
	Typical	Count	46	231	45	322
		% of Total	9.7%	48.6%	9.5%	67.8%
	High	Count	2	37	31	70
		% of Total	.4%	7.8%	6.5%	14.7%
	Total	Count	82	317	76	475
		% of Total	17.3%	66.7%	16.0%	100.0%

In mathematics, in keeping with findings from overall frequencies of each category for each year noted above, the proportions of schools classified as consistently 'typical' over each consecutive-year comparison (29.7% from 2009-10 to 2010-11, 28.8% from 2010-11 to 2011-12, and 26.3% from 2011-12 to 2012-13) were smaller than those in literacy, but still larger than any other combination of categories. Proportions of schools consistently classified as 'high' (17.9% from 2009-10 to 2010-11, 18% from 2010-11 to 2011-12, and 16.6% from 2011-12 to 2012-13) or 'low' (18.5% from 2009-10 to 2010-11, 18.5% from 2010-11 to 2011-12, and 16.2% from 2011-12 to 2012-13) over each one-year period were more than double the corresponding proportions identified in the analyses of schools' literacy performance. Taken together with the correlations reported above to reflect year-to-year stability of school effects and the greater overall variation at the school level in mathematics compared to literacy, it is not especially surprising that a larger proportion of schools remained 'high' or 'low' performing from one year to the next. Again, the smallest proportions of schools were those moving from 'low' to 'high' (0.9% from 2009-10 to 2010-11, 2.2% from 2010-11 to 2011-12, and 0.8% from 2011-12 to 2012-13) and from 'high' to 'low' (0.9% from 2009-10 to 2010-11, 1.1% from 2010-11 to 2011-12, and 1.3% from 2011-12 to 2012-13). The proportions of schools that were, in effect, improving by one category based on their CVA effects from 'low' to 'typical' (8% from 2009-10 to 2010-11, 7.3% from 2010-11 to 2011-12, and 12.6% from 2011-12 to 2012-13) and from 'typical' to 'high' (7.1% from 2009-10 to 2010-11, 7.5% from 2010-11 to 2011-12, and 7.6% from 2011-12 to 2012-13) were not dramatically different from those seen for literacy as reported above. Those essentially declining by one category from 'high' to 'typical' (8.9% from 2009-10 to 2010-11, 6.4% from 2010-11 to 2011-12, and 9.9% from

2011-12 to 2012-13) and from 'typical' to 'low' (8% from 2009-10 to 2010-11, 10.3% from 2010-11 to 2011-12, and 8.6% from 2011-12 to 2012-13) were also quite similar to the corresponding proportions in literacy.

Table 37 Cross-tabulation of school performance categories, 2009-10 to 2010-11 (Mathematics)

Mathematics			2010-11			
			Low	Typical	High	Total
2009-10	Low	Count	83	36	4	123
		% of Total	18.5%	8.0%	.9%	27.5%
	Typical	Count	36	133	32	201
		% of Total	8.0%	29.7%	7.1%	44.9%
	High	Count	4	40	80	124
		% of Total	.9%	8.9%	17.9%	27.7%
	Total	Count	123	209	116	448
		% of Total	27.5%	46.7%	25.9%	100.0%

Table 38 Cross-tabulation of school performance categories, 2010-11 to 2011-12 (Mathematics)

Mathematics			2011-12			
			Low	Typical	High	Total
2010-11	Low	Count	84	33	10	127
		% of Total	18.5%	7.3%	2.2%	27.9%
	Typical	Count	47	131	34	212
		% of Total	10.3%	28.8%	7.5%	46.6%
	High	Count	5	29	82	116
		% of Total	1.1%	6.4%	18.0%	25.5%
	Total	Count	136	193	126	455
		% of Total	29.9%	42.4%	27.7%	100.0%

Table 39 Cross-tabulation of school performance categories, 2011-12 to 2012-13 (Mathematics)

Mathematics			2012-13			
			Low	Typical	High	Total
2011-12	Low	Count	77	60	4	141
		% of Total	16.2%	12.6%	.8%	29.7%
	Typical	Count	41	125	36	202
		% of Total	8.6%	26.3%	7.6%	42.5%
	High	Count	6	47	79	132
		% of Total	1.3%	9.9%	16.6%	27.8%
	Total	Count	124	232	119	475
		% of Total	26.1%	48.8%	25.1%	100.0%

The use of standardised scores for the analyses discussed in this thesis is worth noting with regard to the interpretation of these results. In essence, this means that comparisons of individual students' scores, and consequently measures of schools' performance, comprise what might be described as a "zero sum game" in which school performance is inherently a relative -- rather than absolute -- measure.

7.6.2 Patterns over two-year time spans

Following the same procedures as above, patterns of performance were investigated across two-year time spans from 2009-10 to 2011-12 and from 2010-11 to 2012-13. Results of cross-tabulations for these comparisons are given in Tables 40 and 41 for literacy, and Tables 42 and 43 for mathematics. Relative to the one-year comparisons given above, two-year patterns in literacy are quite similar in terms of the proportions of schools that were consistently 'low', 'typical' or 'high', improving or declining by one category, or improving or declining by two categories. The same was true for mathematics. Overall, there were no dramatic differences between one-year and two-year patterns of categorised CVA school effects.

Table 40 Cross-tabulation of school performance categories, 2009-10 to 2011-12 (Literacy)

Literacy			2011-12			
			Low	Typical	High	Total
2009-10	Low	Count	38	34	4	76
		% of Total	8.7%	7.8%	.9%	17.4%
	Typical	Count	40	219	31	290
		% of Total	9.2%	50.2%	7.1%	66.5%
	High	Count	1	42	27	70
		% of Total	.2%	9.6%	6.2%	16.1%
	Total	Count	79	295	62	436
		% of Total	18.1%	67.7%	14.2%	100.0%

Table 41 Cross-tabulation of school performance categories, 2010-11 to 2012-13 (Literacy)

Literacy			2012-13			
			Low	Typical	High	Total
2010-11	Low	Count	34	40	5	79
		% of Total	7.5%	8.8%	1.1%	17.4%
	Typical	Count	42	220	39	301
		% of Total	9.3%	48.6%	8.6%	66.4%
	High	Count	4	43	26	73
		% of Total	.9%	9.5%	5.7%	16.1%
	Total	Count	80	303	70	453
		% of Total	17.7%	66.9%	15.5%	100.0%

Table 42 Cross-tabulation of school performance categories, 2009-10 to 2011-12 (Mathematics)

Mathematics			2011-12			
			Low	Typical	High	Total
2009-10	Low	Count	69	43	7	119
		% of Total	15.8%	9.9%	1.6%	27.3%
	Typical	Count	54	103	38	195
		% of Total	12.4%	23.6%	8.7%	44.7%
	High	Count	8	37	77	122
		% of Total	1.8%	8.5%	17.7%	28.0%
	Total	Count	131	183	122	436
		% of Total	30.0%	42.0%	28.0%	100.0%

Table 43 Cross-tabulation of school performance categories, 2010-11 to 2012-13 (Mathematics)

Mathematics			2012-13			
			Low	Typical	High	Total
2010-11	Low	Count	70	50	6	126
		% of Total	15.5%	11.0%	1.3%	27.8%
	Typical	Count	46	120	45	211
		% of Total	10.2%	26.5%	9.9%	46.6%
	High	Count	5	53	58	116
		% of Total	1.1%	11.7%	12.8%	25.6%
	Total	Count	121	223	109	453
		% of Total	26.7%	49.2%	24.1%	100.0%

7.6.3 Patterns over a three-year time span

Again following the same strategy as above, school performance categories were cross-tabulated, this time covering a three-year gap from 2009-10 to 2012-13. The results for literacy and mathematics are given in Tables 44 and 45, below.

Other than slightly lower proportions of schools classified as 'high', 'low' or 'typical' in both 2009-10 and 2012-13, and correspondingly slightly higher proportions of schools showing improvement or decline, the three-year comparisons are not very different from those reported above for one- and two-year comparisons. On the whole, this reinforces the idea that school effects are relatively stable over short time periods from year to year, as are patterns of school effects across the school system as a whole.

Table 44 Cross-tabulation of school performance categories, 2009-10 to 2012-13 (Literacy)

Literacy			2012-13			
			Low	Typical	High	Total
2009-10	Low	Count	25	48	3	76
		% of Total	5.8%	11.1%	.7%	17.5%
	Typical	Count	45	201	42	288
		% of Total	10.4%	46.3%	9.7%	66.4%
	High	Count	7	39	24	70
		% of Total	1.6%	9.0%	5.5%	16.1%
	Total	Count	77	288	69	434
		% of Total	17.7%	66.4%	15.9%	100.0%

Table 45 Cross-tabulation of school performance categories, 2009-10 to 2012-13 (Mathematics)

Mathematics			2012-13			
			Low	Typical	High	Total
2009-10	Low	Count	64	45	9	118
		% of Total	14.7%	10.4%	2.1%	27.2%
	Typical	Count	39	116	39	194
		% of Total	9.0%	26.7%	9.0%	44.7%
	High	Count	14	50	58	122
		% of Total	3.2%	11.5%	13.4%	28.1%
	Total	Count	117	211	106	434
		% of Total	27.0%	48.6%	24.4%	100.0%

Ultimately, there was no clear relationship between specific CFN affiliations and particular performance patterns; cross-tabulating patterns of improvement, decline, and consistently

high, typical and low performance showed no notable clusters of schools with similar patterns that were included in certain networks. An investigation of possible relationships between the various performance patterns explored above and network affiliation involves an additional consideration. That is, as schools were able to change affiliation annually, it was also worth investigating possible relationships between whether or not a school kept the same affiliation over a period of time and that school's pattern of performance. Tables 46 through 51 show cross-tabulations including percentages of row totals (representing each performance pattern including consistently low, declining, consistently typical, improving, or consistently high) for two- and three-year time spans in literacy and math. It is apparent from the results shown in these tables that for both subjects and over all of the two- and three-year spans considered, a disproportionately high proportion of schools classified as 'high' at the beginning and end of a particular two or three year period are those that kept the same CFN affiliation. This does not support conclusions about causation, but a possible reason for this apparent relationship between keeping a consistent affiliation and maintaining high CVA school effects over time might have been that highly effective schools had no pressing motivation to change affiliation or to be dissatisfied with whatever support was available.

Table 46 Cross-tabulation of school performance patterns and network affiliation change, 2009-10 to 2011-12 (Literacy)

Literacy 2009-10 to 2011-12		Changed affiliation	Kept affiliation	Total
Low	Count	14	24	38
	Row %	36.8%	63.2%	100.0%
Declining	Count	25	58	83
	Row %	30.1%	69.9%	100.0%
Typical	Count	63	156	219
	Row %	28.8%	71.2%	100.0%
Improving	Count	18	51	69
	Row %	26.1%	73.9%	100.0%
High	Count	4	23	27
	Row %	14.8%	85.2%	100.0%
Total	Count	124	312	436
	Row %	28.4%	71.6%	100.0%

Table 47 Cross-tabulation of school performance patterns and network affiliation change, 2009-10 to 2011-12 (Mathematics)

Mathematics 2009-10 to 2011-12		Changed affiliation	Kept affiliation	Total
Low	Count	30	39	69
	Row %	43.5%	56.5%	100.0%
Declining	Count	25	74	99
	Row %	25.3%	74.7%	100.0%
Typical	Count	28	75	103
	Row %	27.2%	72.8%	100.0%
Improving	Count	27	61	88
	Row %	30.7%	69.3%	100.0%
High	Count	14	63	77
	Row %	18.2%	81.8%	100.0%
Total	Count	124	312	436
	Row %	28.4%	71.6%	100.0%

Table 48 Cross-tabulation of school performance patterns and network affiliation change, 2010-11 to 2012-13 (Literacy)

Literacy 2010-11 to 2012-13		Changed affiliation	Kept affiliation	Total
Low	Count	6	28	34
	Row %	17.6%	82.4%	100.0%
Declining	Count	12	77	89
	Row %	13.5%	86.5%	100.0%
Typical	Count	38	182	220
	Row %	17.3%	82.7%	100.0%
Improving	Count	11	73	84
	Row %	13.1%	86.9%	100.0%
High	Count	0	26	26
	Row %	0.0%	100.0%	100.0%
Total	Count	67	386	453
	Row %	14.8%	85.2%	100.0%

Table 49 Cross-tabulation of school performance patterns and network affiliation change, 2010-11 to 2012-13 (Mathematics)

Mathematics 2010-11 to 2012-13		Changed affiliation	Kept affiliation	Total
Low	Count	6	28	34
	Row %	17.6%	82.4%	100.0%
Declining	Count	12	77	89
	Row %	13.5%	86.5%	100.0%
Typical	Count	38	182	220
	Row %	17.3%	82.7%	100.0%
Improving	Count	11	73	84
	Row %	13.1%	86.9%	100.0%
High	Count	0	26	26
	Row %	0.0%	100.0%	100.0%
Total	Count	67	386	453
	Row %	14.8%	85.2%	100.0%

Table 50 Cross-tabulation of school performance patterns and network affiliation change, 2009-10 to 2012-13 (Literacy)

Literacy 2009-10 to 2012-13		Changed affiliation	Kept affiliation	Total
Low	Count	6	28	34
	Row %	17.6%	82.4%	100.0%
Declining	Count	12	77	89
	Row %	13.5%	86.5%	100.0%
Typical	Count	38	182	220
	Row %	17.3%	82.7%	100.0%
Improving	Count	11	73	84
	Row %	13.1%	86.9%	100.0%
High	Count	0	26	26
	Row %	0.0%	100.0%	100.0%
Total	Count	67	386	453
	Row %	14.8%	85.2%	100.0%

Table 51 Cross-tabulation of school performance patterns and network affiliation change, 2009-10 to 2012-13 (Mathematics)

Mathematics 2009-10 to 2012-13		Changed affiliation	Kept affiliation	Total
Low	Count	6	28	34
	Row %	17.6%	82.4%	100.0%
Declining	Count	12	77	89
	Row %	13.5%	86.5%	100.0%
Typical	Count	38	182	220
	Row %	17.3%	82.7%	100.0%
Improving	Count	11	73	84
	Row %	13.1%	86.9%	100.0%
High	Count	0	26	26
	Row %	0.0%	100.0%	100.0%
Total	Count	67	386	453
	Row %	14.8%	85.2%	100.0%

7.7 CHAPTER SUMMARY

Findings from this chapter further extend those from Chapter 5 and again support the conclusion that there is little evidence to suggest any significant overall 'network effect' of CFNs on student achievement in literacy and mathematics during the individual years studied or across the three years for which performance data were available. In relation to the NYCDOE claim introduced at the beginning of this thesis that these networks were supposed to help raise achievement and support school improvement, these findings provide little empirical support.

Correlations between contextualised value added school effects for literacy and mathematics across all of the years considered, from the year prior to full implementation of the network structure (2009-10) to the final year for which data were available at the time of analysis (2012-13), were strong for empty models and moderate for CVA models. The same was true for correlations reflecting consistency between subject areas. That is, unadjusted (gross) school effects were extremely stable and consistent, but considerably less so after accounting for student background and prior achievement as well as a measure of school context (school mean intake achievement). The fact that school effects are fairly stable overall helps to justify the use of 2012 results to inform the sampling of schools for qualitative study, while the finding that CVA residual estimates of school effects are less

stable than gross school effects underscores the importance of taking change over time into account in analysing school performance and controlling for differences in student and school intake characteristics. This was addressed in the analyses in the present chapter, but also explored in the qualitative findings presented in Chapter 6 based on interviewees' accounts of changes in school-based and system-wide policy and practice (for example, changes in priorities for school improvement and support, and changes in pedagogical emphasis as a consequence of the rollout of a new learning standards initiative, the Common Core State Standards). Additionally, the finding that CVA school effects are less consistent than gross effects adds justification for the use of more than one subject area outcome in the present study and the control for student intake.

Finally, an examination of patterns of school performance over time based on categories of 'high', 'low' and 'typical' school residuals from CVA models showed that while there were no clear patterns of improvement or decline in school performance (as measured by CVA residual estimates of school effects) associated with specific CFN affiliations, schools maintaining high performance according to these measures in each subject area tended to be those that did not change CFN affiliation. This does not offer evidence of a relationship between CFNs and school improvement. It does, however, suggest the possibility that either effective schools made better use of the network support structure, or were less likely to take advantage of (or require) the option to change affiliations annually, a notion that is supported by the qualitative findings in the previous chapter. It may also have been the case that more effective schools made more appropriate choices of which network to affiliate with at the start of the policy.

8 DISCUSSION AND CONCLUSIONS

8.1 CHAPTER GUIDE

This chapter is intended to:

- Draw together findings from the quantitative and qualitative strands of the project as presented in previous chapters
- Discuss and interpret the findings in the context of the relevant academic literature
- Provide an analysis of the strengths and limitations of the study and the original contribution to knowledge
- Identify directions for future research
- Discuss implications for policy

8.2 CHAPTER INTRODUCTION

The overarching aim of this research was to address a set of claims about Children First Networks in the NYC DOE. In the description of the stated purpose that the Children First Networks were intended to achieve , it was claimed specifically that they would facilitate *collaboration and learning* across schools, and that they would *help schools to raise achievement*. Findings reported in the previous chapters served to answer research questions aligned with this overarching aim. This concluding chapter seeks to interpret those findings in relation to each other, linking the qualitative and quantitative elements of this mixed methods study. In addition, the findings are discussed in relation to the relevant literature on the role of school networks in relation to student achievement and school improvement more generally. The discussion also seeks to draw links between the findings and the context in which the Children First Networks were introduced, and to explore implications for policy and practice as well as directions for further research.

The first section of this chapter, below, summarises main findings from each of the preceding chapters and discusses conclusions integrating and synthesising evidence from both qualitative and quantitative strands of the study to establish the extent to which findings from the different strands (qualitative and quantitative) complement, challenge or extend each other in order to establish an overall narrative of the 'story' told by this research. The findings are then discussed in the context of relevant literature on networks

and collaboration in education, highlighting the original contributions made by the present study as well as the gaps addressed in the literature. Overall strengths and limitations of the study are revisited, and future directions for research are identified and explained. The chapter concludes with an interpretation of findings with respect to the particular context and policy changes within that context since the project began, and a discussion of implications for policy in the local context and more broadly (with some caution regarding generalizability).

8.3 INTEGRATED CONCLUSIONS FROM QUANTITATIVE AND QUALITATIVE EVIDENCE

Findings from the quantitative strand of this study showed that there was no overall effect of networks, as there was no evidence of significant variation between CFNs in terms of student achievement in either literacy or mathematics. This was true for all years considered. Gross school effects were extremely stable over time in both literacy and mathematics, and contextualised value added school effects were at least moderately stable, while the overall proportions of high-, typically- and low-performing schools remained similar over time. Taking all of these findings into consideration, there is little evidence to suggest that these networks helped to raise student achievement.

There was a lack of clarity amongst school-based practitioners regarding the roles and purposes of CFNs, and these networks were not seen as change drivers or as mechanisms for school improvement. Although some network staff explicitly described school improvement planning processes, CFNs were mainly viewed by school-based staff as support organisations that disseminated information, helped principals to maintain awareness of and compliance with current regulations, and acted as a resource for operational support and troubleshooting with regard to immediate issues and problems in schools. This overall lack of agreement and clarity about the role and impact of CFNs with regard to school improvement both elaborates on and may help to explain the lack of an overall network effect on student achievement found via secondary data analysis.

Although facilitating collaboration and learning across schools was part of the description of CFNs on the NYCDOE website, findings from interviews in this study suggest that collaboration and interaction between schools was fairly minimal and limited, for the most part, to interactions between principals. Teachers were mostly not engaged with CFNs at all, and few had engaged with other schools through their schools' networks. Although the

quantitative results do not directly address collaboration, the combination of findings, namely that there was no overall network effect and little evidence of inter-school collaboration, suggests that CFNs were not functioning as school improvement mechanisms in practice in the ways they were intended to (in relation to the two stated purposes noted above (promoting collaboration and learning across schools and raising student achievement) nor, in fact, as meaningful networks of schools. Rather, the evidence produced by this research suggests that each CFN functioned mainly as a support-providing organisational entity associated with, but not necessarily forming substantive links between, a group of schools. The interviews conducted as part of the qualitative strand provided little evidence to suggest that CFNs were perceived by school-based staff (principals and teachers) to have promoted between-school collaboration, professional development or learning across schools.

Findings from interviews provide some insight as to why CFNs seem not to have functioned as either school improvement mechanisms nor, in fact, as facilitators of collaboration between schools. The lack of well-defined and agreed-upon goals of CFN interaction and activity may not have provided sufficient incentive to drive collaborative engagement between schools, nor adequate focus to foster meaningful school improvement. Furthermore, although being able to change affiliation annually may have engendered school autonomy, in cases where schools spent little time in a particular network there might simply not have been enough time to initiate and maintain meaningful educational change. This may also have inhibited the fostering of trust and relationships that previous research suggests is crucial to successful efforts in both school improvement and school-to-school collaboration (Muijs et al., 2011). The question of trust extends beyond individual networks to trust in the system as a whole; the sense that too many changes were happening simultaneously and rapidly may have limited the perceived value on the part of school-based staff – and consequently the impact – of network participation. The fact that teachers rarely participated directly in CFN activities, or expressed a lack of awareness of the roles and purposes of CFNs, suggested that either the design or implementation of network engagement and participation inherently did not embed professional learning communities across groups of schools. Lastly, the challenges of networking and collaboration in challenging circumstances have been established by previous literature in the field (Chapman, 2013); in New York City, the combination of accountability pressures (including, in some cases, the threat of closure) and multiple

policy changes requiring schools to adapt and comply with new requirements meant that principals and teachers may have had more pressing priorities than networking and collaboration, and maintaining compliance may well have taken precedence over concerted school improvement planning and implementation at both the school and network levels.

Having addressed what these networks were not, it is also worth considering what they were. One of the original drivers of this research was a lack of clarity and transparency surrounding the purpose, form and function of CFNs. Revisiting the typology for networking and collaboration in education proposed by Muijs et al. (2011) provides a useful framework for addressing this underlying aim. The qualitative findings help to illuminate where CFNs were situated with regard to the dimensions of the typology, each of which is discussed in detail below.

Goals and activities: While network staff tended to identify school improvement as a key goal of network activities and interactions, school-based staff largely described activities more closely related to broadening opportunities (including making connections with external organisations and securing funding) and sharing resources (mainly in terms of information disseminated by CFNs). This lack of agreement across stakeholder groups has the potential to contribute to theory development; while the authors of the typology characterised goals and activities as relating to substantive focus, time scale and connections to theories of networking, findings presented in this thesis suggest that in fact alignment (between the goals of various stakeholders in a network, and between perceived goals and activities in practice) is also an important aspect of this dimension.

Voluntarism or coercion: CFNs inherently occupied a complex multiplicity of positions on this continuum. Purely based on relevant NYCDOE policy, schools were both required to belong to a network and allowed to opt in and out of a specific network annually. On a more nuanced level, there were variations in the extent of voluntarism dependent upon a principal's level of experience and perceived empowerment, particularly with regard to navigating CFN support, opting in or out of specific activities, and initiating change in a network.

Power relations: Overtly, schools were not directly accountable to CFNs, and were responsible for rating their CFNs annually. This seems to have defined a sort of provider-client relationship in some cases, so that power was largely located at the level of the principal. The dynamics of power relations in the CFN structure were complicated,

however, by the fact that – as some of the interview data indicated – CFNs were at least in some instances staffed by individuals who had worked in positions of authority in the NYCDOE previously, so that their perceived power and authority may have exceeded the bounds of their explicit job roles.

Network density: Maximum network density can be described as the circumstance in which "everyone is connected to everyone else" (Muijs et al., 2011, p.42). CFNs, by contrast, were characterised by relatively minimal density, with principals mainly connecting to certain network staff members and to other principals. With respect to depth of involvement, another aspect of network density, qualitative findings demonstrate that CFN participation mainly involved meetings between participants rather than deeper involvement in collaborative joint activities.

External involvement: This was a key feature of the CFN structure because of the existence and roles of the organisational entity central to each network. While the nature and features of this external involvement varied (with PSO networks associated with university or charity organisations, and non-PSO networks associated with the central NYCDOE), it was clear that CFN participation was defined by the involvement of network personnel operating beyond the individual affiliated schools. Some of the qualitative evidence indicated further external involvement in instances where CFNs had facilitated links between schools and additional external organisations for funding or specific forms of support and partnership.

Time frame: The policy allowing schools to re-affiliate annually meant that there was a variable time frame associated with CFNs. Some schools maintained a long-term association with the same network since the CFN structure was introduced, while others had been with their current CFN (at the time of data collection) for less than a year.

Geographical spread: Principal interview findings suggested that there was a great deal of variation with regard to geographical spread in CFNs. Some groups of schools affiliated with a single CFN were scattered across all five boroughs of New York City (at a considerable distance from one another), while there was at least one example in the sample for this study of a group of schools within one geographical district that had intentionally chosen the same CFN affiliation.

Density of schools: CFNs typically included a number of affiliated schools between 24-34 as defined by local policy. However, participant's accounts suggested that in fact the

density of these networks in terms of the number of schools actively engaging with network activities was sometimes much smaller.

Within- and between-school networking: For the most part, there was little evidence of between-school networking in CFNs, with the exception of monthly principals' meeting. There were, however, opportunities for further interaction between schools in which some principals had chosen not to engage themselves or their staff. In terms of within-school networking, there was some evidence to suggest that information disseminated by CFNs formed the basis for within-school activities involving principals and teachers (for example, with regard to supporting the implementation of a curriculum to address new state standards).

Diffuseness: The extent to which CFNs consisted of a fixed and finite versus loose and shifting group of actors was not clear cut. In one sense, there was (within each year) a clear-cut group of schools affiliated with a given CFN. However, these affiliations also depended on a looser cluster of informal professional networks that drove network choice. Additionally, the group of individuals actively engaging with a particular CFN was not set; in most cases, principals and network staff were the most actively involved, but there was some limited evidence of (often more sporadic) teacher involvement.

Formalisation: In one sense, the CFN structure was highly formalised. School affiliations with CFNs were clearly delineated, and each CFN included a defined central organisational entity with specific roles for network personnel. On the other hand, there was strong qualitative evidence of the centrality of trust and personal relationships in driving formal CFN affiliations and interactions.

While the description of where CFNs sit within this typology is by no means straightforward with respect to many of the various dimensions above, this description still provides a much more thorough illustration of what CFNs were and what they did than was available when this research began.

8.4 CONTRIBUTION TO THE RELEVANT LITERATURE

To address the contribution of the present study, the organisation of this section moves from specific to general. It first addresses how this thesis contributes to the literature specifically on CFNs in New York City, and then moves on to expand the discussion to considerations of how the thesis contributes to the broader literature on networks and collaboration as well as to the educational effectiveness and school improvement research knowledge base.

8.4.1 Contribution to the literature on CFNs

As noted previously in the Introduction (Chapter 1) and Literature Review (Chapter 3), there have been several other studies specifically of CFNs in New York City, the majority of which were published while this study was underway. This section focuses on how the present study adds to this context-specific knowledge base, as well as some of the ways in which the findings converge with or, by contrast, diverge from, findings and conclusions drawn from these other studies.

Substantive contributions

The assessment conducted by the Parthenon Group (2013) concluded that CFNs were "an improvement over earlier organizational designs" (p.4), that "choice introduces continuous improvement" (p.5), and that that "schools prefer to collaborate with a much broader range of their peers than the old district or regional structures allowed" (p.4). Findings from the present study, on the other hand, suggest that not much collaboration was taking place between schools in CFNs, and there was little evidence to support claims of measurable improvement associated with networks. Moreover, the evidence suggested that in some cases schools tended to link with other neighbourhood schools and that the earlier district structure was replicated in the network's membership; to an even greater extent, qualitative findings indicate that network affiliation is often related to principals' pre-established informal professional networks, similarly indicating that network groupings may have replicated pre-existing configurations and connections within the school system. As the Parthenon report lacks clear articulation of the specific approaches to analysis that were adopted that led to the conclusions stated above, it is difficult to establish with certainty the reasons underlying these discrepancies. One reason for differing conclusions about collaboration may be fairly simple, however. In elaborating the point about schools preferring to collaborate with a wider variety of other schools, the report notes that networks are comprised of schools from a variety of geographical districts and boroughs (Ibid.). However, as qualitative findings from the present study demonstrate, just because a school was affiliated with a CFN did not necessarily mean that it was engaged in collaboration with other affiliated schools. With regard to continuous improvement, the Parthenon report suggests that larger numbers of schools tended to join more effective networks, with effectiveness framed in terms of DOE ratings of networks, Progress Reports of other schools in a particular network, and high principal satisfaction. Principal

accounts in the present study did not indicate that these were factors in their network choices, with the exception of principal satisfaction (but couched in terms of word of mouth rather than the satisfaction surveys referred to in the aforementioned report). Furthermore, in a school system in which accountability relies heavily on performance on state assessments, the lack of evidence of an association between CFNs and student achievement on these assessments raises questions about the strong claim made about continuous improvement. On the other hand, the Parthenon group's report does suggest that support for the lowest-performing schools was a challenge in the CFN structure, recommending more prescriptive support for these schools. Findings from the research reported in this thesis, not least the disproportionate number of low-performing schools that changed networks over time, lend empirical evidence to support to this conclusion. Changing network membership was seen, by the few principals in the study who had done so, as a strategy that might be necessary to improve performance or avoid failure.

Another study conducted by the Annenberg Institute for School Reform at Brown University (Fruchter et al., 2015) focused on demographics and academic outcomes related to CFN affiliation. This study found that there was less segregation based on demographic characteristics in CFNs than in geographical district groupings of schools, but found little evidence of an association between CFN affiliation and academic outcomes. Multilevel modelling was used to investigate this potential association. However, variation between students was not accounted for, as the models for which results were reported were based only on schools nested within networks, thus missing the important student level. While the present study lends additional confirmation to the conclusion that networks have little association with academic achievement, a much higher proportion of total variation in academic outcomes before accounting for demographic variables (12-14% in literacy and 11-18% in mathematics) was found to lie between networks in the models used in the study by Fruchter et al. (2015) than was the case in the present study (0.5-0.8% and 0.4% in mathematics for the same years). The discrepancy in the apparent size and statistical significance of network effects between the Annenberg Institute's report and the findings reported in this thesis are likely a consequence of the former failing to account for variation between students, which was mistakenly attributed to higher (school and network) levels. It also prevented the proper modelling of the role of intake differences between individuals. As such, the comparisons of model structure reported in Chapters 5 and 7 are an important new contribution to knowledge on the links between networks and

variation student achievement (attainment and progress) made by this DPhil study to the existing scholarly literature. Beyond the specific context, this also underscores the importance of making such comparisons in the study of network structures in other settings, as the level or classification structure of multilevel models can substantially affect conclusions about the statistical significance of variation between units (e.g. networks) at a particular level (Luyten & Sammons, 2010).

Research on New York City's CFNs conducted by Wohlstetter, Smith and Gallagher (2013) used purely qualitative approaches to explore the network structure, its processes and interactions through the lens of organisational learning theory. Some of the findings from the present study accord with those of Wohlstetter and colleagues. In particular, both studies found that there were a variety of rationales for network choice, including principals' pre-existing social networks, geography , and shared educational values. The centrality of the role of principals in shaping network structure and CFN support for schools was also a convergent finding from both studies. Regarding the extent and nature of collaboration amongst schools affiliated with a given CFN, the study by Wohlstetter et al. (Ibid.) suggested that the network structure provided opportunities for collaboration but that the extent to which collaboration took place was often quite variable. The evidence from interviews conducted for that research point to some examples of collaboration that were tightly defined around specific goals and organised around schools' or principals' areas of interest or concern. However, the present study found only limited evidence of such goal- or interest-oriented collaboration, although there was evidence of interaction mainly in the form of principals' meetings. One reason for this apparent difference in findings is likely to be due to the composition of the samples of individuals interviewed for the two studies; while Wohlstetter et al. conducted 21 interviews, 3 of which were with principals and the rest of which were with individuals at higher levels of the network structure and NYCDOE, the interview participants for the present study included a majority of school-based staff (principals and teachers) and only a few network staff. Findings reported in Chapter 6 of this thesis indicate a lack of alignment in the perspectives of these different stakeholder groups with regard to network roles as well as the extent to which networks facilitated collaboration and improvement, making the discrepancy in findings noted above less than surprising. Ultimately, this should not be viewed as an outright contradiction of the research by Wohlstetter and colleagues, rather, the divergence itself highlights the contribution the present DPhil study makes to the

relevant knowledge base by offering evidence based on the perspectives of different stakeholder groups.

An additional study by Wohlstetter, Houston, and Buck (2015) focused on the role of CFNs in Common Core[16] implementation in New York City. While the present study was not intended to focus on Common Core rollout, the data collection occurred during roughly the same time period for both studies, and findings show some overlap with regard to teachers' lack of engagement with CFNs and the "hub and spoke model" (Ibid., p.106) in which CFN organisational entities interacted with schools but schools rarely interacted with each other. While Wohlstetter et al. exclusively interviewed network staff, the present study makes a unique contribution by supporting these findings with evidence from teacher interviews and evidence from a larger number of principals.

Methodological contribution

Of the research presented above, all employed exclusively quantitative or qualitative approaches, with the exception of the Parthenon Group (2013) report, for which the methods were not reported with specificity and it was somewhat unclear what quantitative analysis had been conducted. One of the unique features of this doctoral research, then, is the deliberate choice of a mixed methods design to study CFNs and to allow both qualitative and quantitative evidence to be collected, analysed and linked. This constitutes an original contribution to the relevant knowledge base by attending to both academic outcomes and their relationships (or lack thereof) to CFN affiliation and the CFN structure as a whole, and the perspectives and experiences of those participating in these networks.

8.4.2 Contribution to the literature on networks and collaboration in education

The present study makes original contributions to the existing research knowledge base on inter-school networks and collaboration both substantively and methodologically.

Substantive contributions

Substantive contributions to the knowledge base on networks and collaboration in education can be divided into three categories:

- Evidence regarding student achievement

[16] Note: "Common Core" refers to a set of new standards and learning goals implemented across New York State in recent years, leading to substantial changes in curriculum and instruction as well as a change in the state assessments in 2012-13.

- Evidence regarding leadership roles in schools and networks
- Evidence regarding engagement and participation in collaboration among schools

Impact on students, particularly on academic outcomes, has been a common focus in research on school networks and related collaborative initiatives. Existing empirical evidence of the effects of networking and collaboration prior to this study is, at best, limited and mixed. While a number of studies do conclude that networks can affect student achievement (e.g. Chapman & Muijs, 2014; Greenberg, 1996; Hutchings et al., 2012; Montgomery, 2001; Zetlin, MacLeod, & Michener, 1998), the structures for collaboration between schools vary so widely that these conclusions cannot be reasonably generalised to networks more broadly. A further complication arises from the fact that achievement outcomes, too, are varied and not always measured directly; some studies claiming impact on students draw on teacher surveys to assess impact on student learning (Bell et al., 2006), while others define student achievement in terms of narrowly-defined, highly specific outcomes (e.g. Bielefeldt, Moursund, Underwood, & Underwood, 1999). Additionally, the use of only school-level data forms the basis for conclusions about student achievement in many studies, including those finding little or no effect of networking (e.g. Sammons, Mujtaba, Earl, & Gu, 2007). The findings of this DPhil study, in particular the lack of a network effect, contributes to the existing knowledge base on the effects of networks on student achievement in that 1) these findings draw on directly-measured achievement at the student level, 2) multiple achievement outcome measures were used, specifically literacy and mathematics (both of which are commonly accepted as core subject areas), and 3) multilevel modelling allowed for analyses to account for variation between students and avoided mistakenly attributing upwardly-biased proportions of variation to higher-level school and network units.

The roles and perspectives of school leaders has also formed a common focus or dominant thematic strand in findings from research on inter-school collaboration and networking. The bulk of this research has been conducted in the UK in a variety of different network and partnership structures (e.g. Coleman, 2012; Hopkins & Higham, 2007; Lindsay et al., 2007). Findings from the present study regarding the centrality of the role of principals in shaping network support in a fully-networked school system offer an original contribution to the body of knowledge regarding school leadership within school networks, as this evidence offers insight from a context and network structure not previously studied with an emphasis on school leaders' perspectives.

Finally, this research makes an original substantive contribution to the existing literature on participation and collaboration in networks based findings of a lack of alignment in stakeholder perceptions of the extent to which collaboration was taking place, and a lack of teacher participation in network activities supported by evidence from teacher perspectives. Previous research has explored the extent and nature of engagement and collaboration in networks, but many studies engaging in this line of inquiry concern network structures and initiatives with more clearly-defined processes or goals for linking teachers across school boundaries (e.g. Stoll, 2015). The present study makes a unique contribution in that 1) it finds minimal teacher engagement with networks, and even minimal awareness of collaboration as a stated purpose of networks in New York city schools, and 2) these findings are not based on the accounts of school or system leaders, but draw directly on interviews with teachers themselves.

Methodological contribution

In terms of the methodological approach, the use of mixed methods involving advanced quantitative approaches such as multilevel modelling is relatively rare in the study of the effects of networks of schools. The use of mixed methods, however, is not uncommon, as research involving case study designs with some combination of interview and survey approaches are prevalent in the field (Atkinson, Springate, Johnson, & Halsey, 2007). The use of multilevel modelling or other sophisticated statistical approaches is also not without precedent in the study of networks (e.g. Chapman & Muijs, 2014), and the combination of multilevel modelling of secondary data and interviews with school leaders is rare but not non-existent in previous research (e.g. Chapman & Muijs, 2013). However, this research was conducted in England and no equivalent, to the author's knowledge, has been conducted in other contexts including the US. That said, the combination of what might be termed multilevel qualitative approaches (i.e. involving stakeholders at various levels of the school system) and multilevel approaches treating networks as a potential level in a fully-networked school system, rather than making comparisons between networked and non-networked schools, is an original methodological contribution of this DPhil study.

8.4.3 Contribution to the literature on educational effectiveness

Previous studies of educational effectiveness have largely focused on the effects of teachers and schools, as well as, to a lesser extent, neighbourhoods and other higher-level units such as countries (e.g. Lenkeit & Caro, 2014). Studies using multilevel modelling

and educational effectiveness approaches to study school networks, however, are rare. The work of Chapman and Muijs (2014) is an exception, comparing federated to non-federated schools in England using multilevel analysis. However, the nature of the research questions and the availability of data on non-federated comparator schools made it appropriate to model federation as a fixed effect. The present study's approach of incorporating network as a level and comparing the role of the network as a third level with that of the district, as well as exploring cross classifications of district and network, may be seen as a more appropriate approach in contexts where no comparator group is available. The Annenberg Institute report (Fruchter et al., 2015) on CFNs mentioned in Section 8.4.1 provides another example of an educational effectiveness approach to studying school networks, but does not fully account for the multilevel structure of the education system at hand by omitting the student level, a major omission which may have biased conclusions about the proportion of variation attributed to networks. Similarly, the lack of the student level in the evaluation of network learning communities in England was highlighted as an important limitation by Sammons, Mujtaba and Earl (Sammons, Mujtaba, et al., 2007).The present study makes an original contribution in its particular approach to applying educational effectiveness research methods to the study of networks, treating networks as a level in the model comparisons conducted for each of the three years considered. This offers a substantive contribution, specifically providing robust findings on the lack of evidence of a statistically significant network effect, as well as a methodological strategy that might be used to study the effects of networks on student outcomes in other settings.

8.5 STRENGTHS AND LIMITATIONS

This research had several strengths worth noting. The findings draw on multiple sources of data, including qualitative and quantitative as well as primary and secondary data. This informed a multidimensional view of CFNs in terms of student and school achievement and the experiences and perspectives of those participating in these networks. Within the quantitative strand alone, the use of multiple outcomes (both literacy and mathematics) provided a fuller picture of student and school achievement in the network structure than would have been possible had only one outcome measure been considered. The analysis of three successive years of data also allowed stability and change in school performance to be explored, although the lack of vertical equating in the assessments used precluded the use of growth curve analysis. Additionally, interviewing stakeholders at different levels of

the network structure including teachers, principals, and network leaders, incorporated multiple perspectives and illuminated issues that might not have arisen if these groups had not been included as participants (for example, teachers' lack of engagement with CFN activities and often their lack of awareness of the CFN structure and its intended purpose). Attending to multiple perspectives in context is a key strength of the study, as this allowed for the possibility that the implementation of the CFN structure differed in practice from the claims this research was designed to address, and highlighted some of the possible reasons for the quantitative findings that emerged. The deliberate linking of the qualitative and quantitative findings provides further insights as it has been possible to establish how far the quantitative and qualitative findings complement, extend or challenge each other. In this way, the integration and synthesis has allowed broader conclusions to be drawn.

Of course, as is the case for any research, this project also had a number of limitations. These included:

- Data issues: While a considerable amount and scope of secondary data was available, both publicly at the school level and by request at the student level, there were several problems inherent to the data available that precluded some of the otherwise preferable approaches to analysis. Missing data was handled by using multilevel multiple imputation, and this helped to alleviate the problem. On the other hand, the state assessments were not vertically equated across grade levels and this unfortunately prevented the use of growth modelling, which would otherwise have been the best choice of statistical modelling approach to examine changes in achievement over time at the student and school levels and this might have provided additional insights into potential effects at the network level. Given the limited variance at the network level identified in models in each year, it is not highly likely that growth curve models would have altered the conclusions about network effects drastically. However, growth curve models have demonstrated larger effects (of teachers and schools, for example) on achievement change than on achievement status in previous educational effectiveness research (Ortega, 2016).
- Access issues: Despite the originally intended approach to sampling for the qualitative case study component of this research, access challenges led to a sample that was, in practice, necessarily a convenience sample. This was a concession to the practical 'messiness' of qualitative research in context.

Ideally, it would have been preferable to have a purposive sample of schools, possibly including selections of schools covering a range of effectiveness within each of a small number of CFNs (drawing on high, low and typical classifications of school performance such as those developed in Chapters 5 and 7).

- Causality: It would be over-claiming to make conclusive causal attributions based on the findings presented in this thesis. More than a few other initiatives and policy changes affected schools over the same time period, and without in effect having a control group to which CFN affiliated schools could be compared in experimental fashion, it is not possible to entirely disentangle the effects of CFNs from the effects of these other changes, pressures and policies in operation at the time in New York schools. The quantitative approaches used are only able to demonstrate association (or lack of association). While qualitative findings help to elaborate on and 'explain' the statistical results and to provide insights that deepen understanding of what networks meant to different stakeholder groups, the qualitative results can inform speculation and assist in the development of tentative explanations that may suggest possible causal relationships, but cannot prove them. For example, the lack of inter-school collaboration as well as teachers' lack of awareness about the roles and puposes of networks and their lack of engagement in network related activities may help to account for and enhance understanding of the lack of any significant network effects identified by the large-scale statistical analyses.
- Generalisability: Findings and conclusions are specific to the context in which this research was conducted. The research focuses on three years of school performance data and relates to the particular contexts of New York City schools. Results have the potential to inform policy in that setting, but generalisations to other settings must be made with caution.

In the interest of reflexivity, it is worth taking note of the fact that this was a doctoral research project. As such, an implicit part of the process involved my own training and development as a researcher. Given the quite ambitious scope of this research, the mixed methods design, and the use of fairly advanced quantitative methods, there was a steep learning curve from the design phase to the point of writing up, including training and self-directed study in numerous aspects of qualitative, quantitative, and mixed methods

research (from philosophy and theory to technical aspects of the various types of software employed for analysis). I would suggest that my being a doctoral student was related to some of the limitations of this research. Specifically, my student status may well have been a factor in the challenges involved in gaining access to schools, and the fact that this was a single-researcher endeavour, along with the need for various types of training, certainly played a role in the time required from start to finish. In particular it meant that only a relatively small number of schools and participants could be included in the interviews for the qualitative strand of the research.

8.6 DIRECTIONS FOR FUTURE RESEARCH

The present study was bounded by a specific set of aims and research questions, and was limited by the practicalities of being conducted by a single researcher within the timeframe of a doctoral course. However, there are a number of valuable directions for future research arising from this DPhil project, some of which would build upon the findings of this particular study about networks in New York, and some of which are inspired by data and access challenges encountered over the course of this project. This is by no means an exhaustive list, as there could certainly be additional possible directions for further inquiry, but the following describe the directions for future research that emerged as some of the most pressing or natural extensions of this work:

1) A large-scale survey approach: Interview methods were used in this research because the research questions merited an exploratory approach; little was known at the outset about the processes and interactions relevant to CFNs in and between schools. A natural extension of this approach would be to follow up the use of interviews by designing questionnaires and administering these in electronic format in order to access a much larger sample of schools and individuals at minimal cost. Results from a questionnaire survey would facilitate additional approaches to analysis, notably social network analysis, which would have the potential to yield further insights about the processes and interactions within networks. It was not practical to include a survey as an additional stage in the design of the present study, mainly because of the time limits of doctoral research, and given that CFNs will no longer be in place after the completion of this study, it will not be possible to use this as a direct follow-up on the same topic in New York City schools based on the network structures studied in this thesis. However, the same approach would

be useful for studying similar networks in other settings, or for studying new structures of collaboration and interaction between schools in New York City currently being piloted (see Section 8.7).

2) Direct observation of network-related activities: This would be another useful extension of the current research. Interviews provided accounts of network-related activities and interactions from various participants' perspectives, but direct observation would have potentially provided additional insight and a form of triangulation with regard to the extent to which findings from observation and interview data converged or diverged. Direct observations were not possible in this study due to a combination of time constraints and access challenges. However, like the survey approach described above, including observations along with interviews could benefit future studies of similar networks elsewhere or new forms of school networks in the same setting.
3) Alternative inter-school collaboration approaches in the same setting: Another logical follow-up would be a study new structures for inter-school collaboration that have emerged more recently in New York City. The advantage of studying such networks and partnerships between schools would be that, in contrast to CFNs, not all schools are required to participate in a particular type of collaboration, so there is more scope for comparison and possibly more support for developing explanations and making claims about the benefits of impacts of different approaches to collaboration. If possible, returning to some of the same schools would potentially facilitate the comparison of processes, interactions and perspectives, as well as outcomes, between CFNs and these new school partnerships and networks. Depending upon the specific goals of particular partnerships or groupings of schools, this line of inquiry might also fit appropriately with the additional consideration of non-academic outcomes that are also relevant to school improvement (e.g. bullying, attendance, or parent involvement, to name a few possibilities).
4) Similar approaches to studying school networks in other settings: The methods used in this study might also provide a useful basis for research designs to yield insights about structures for inter-school collaboration and networking in other settings (academy chains or Teaching School Alliances in England, for example).
5) The notion of alternative outcome measures constitutes yet another direction for future research. Scaled scores for academic outcomes were used in the research

presented in this thesis, but performance levels based on scale score cutoffs are widely used in the New York City school system for accountability purposes, and these also often form the basis of conversations around student performance in schools. While it was beyond the scope of this project, future research from an educational effectiveness perspective comparing different approaches to modelling school effectiveness (e.g. using scaled scores as outcomes and using performance levels from 1='well below proficient' to 4='excelling' as outcomes) would be appropriate to investigate whether the use of differently-defined outcomes for the same assessments might lead to different conclusions about effectiveness at the school level as well as at higher levels of the school system (e.g. networks or collaborative groupings of schools). Moreover, it would be useful to consider outcomes other than student achievement to provide a broader picture of school effectiveness and improvement. In interviews, the goals and priorities for development and improvement mentioned by principals and teachers were not limited to promoting students' academic outcomes, suggesting that it would be worthwhile to investigate effects beyond those on literacy and mathematics achievement. School surveys are administered annually in New York City public schools to teachers, students and parents, and these could provide a rich source of secondary data for such analyses, as they include items relevant to school environment, instruction, leadership, and family and community engagement. These can be seen as important outcomes in their own right, but also in many cases these function as important features of school organisational processes that may support school improvement and better outcomes for students in educational effectiveness models.

8.7 IMPLICATIONS FOR POLICY

Since the beginning of this study, policy changes extremely relevant to the focus of this research were enacted in the New York City Department of Education. In particular, a new set of structures for partnership and collaboration between schools have emerged, and CFNs will have been phased out by the time this thesis has been submitted. Nonetheless, this research has several implications for policy relevant to implementation of school networks in this setting and possibly (though cautiously, in acknowledgement of the context-specificity of the findings) in other settings where policymakers might aim to implement similar networks.

The evident lack of agreement and clarity regarding network role and purpose underpins one of the major implications of the study. CFNs offered some opportunities for collaboration and interaction between schools, but there was not consistent buy-in or participation in these activities on the part of principals and teachers. Not all principals saw a clear incentive for or benefit of these opportunities for collaboration and interaction, and as CFN support offerings were optional, they simply opted not to participate. Overall, there was a lack of defined goals, commitment, and shared understanding about how networks might be expected to support better student outcomes. There was ostensibly no theory underlying their introduction that might have supported these aims. This leads to three recommendations. First, if the purpose of school networks is to foster collaboration, school-based staff need to be fully informed about this purpose if they are to be expected to participate. Second, clear benefits would provide an incentive to participate, and engaging practitioners directly to establish a focus and a set of goals for collaboration would make benefits explicit and relevant to schools. Third, rather than a menu of collaborative opportunities for principals to opt in or out of on a one-off basis, an initial contract between schools and the network organisational entity would address the issue of commitment; this might be combined with the goal-setting described above so that school-based practitioners would have a voice in shaping the contract, which in turn might foster buy-in as well as commitment.

With regard to raising achievement through school networks, the implications for policy are less obvious. Given the relatively weak and equivocal evidence base about the potential of school networks to improve academic outcomes, there is no guarantee that networking schools is necessarily the best approach to support school improvement for the purpose of raising student achievement. However, I would suggest that the above recommendations, while not guaranteed to raise achievement, would at least ensure more direct, sustained engagement and participation of schools in their networks, which would at a minimum increase the possibility that networks might have a measurable effect.

Another implication for policy concerns the option for schools to re-affiliate annually. If school networks are formed for the purpose of school improvement, the fact that low-performing schools were disproportionately represented amongst those that changed networks suggests a fundamental flaw. One year is a short span of time to effect meaningful change in a school; it might be advisable, therefore, for policy establishing

networks of schools to specify a commitment from schools covering a longer time period. This might also make it possible to evaluate network effectiveness more reliably.

Interestingly, new structures for inter-school collaboration mentioned above have already begun to address these recommendations. For example, 'Learning Partners', an initiative entering a piloting phase in 2016-17, involves an emphasis on a specific focus area, and involve a structured relationship between three to four collaborating schools in which one school is identified as having a specific strength while other partner schools have identified the same area as a priority for improvement (NYCDOE, 2016b). Notably, 'Learning Partners' involve a more prescribed structure for collaboration, with monthly inter-visitations and bi-weekly school team meetings. This links with at least two of the recommendations above, namely, a specific purpose and goal for collaboration and a commitment to consistent participation in collaborative activities. It also links with the findings in England by Chapman and Muijs (2014) that note the positive effects of strong federations where weaker schools are linked with a more effective school as cluster leader.

The Middle School Quality Initiative, which began in 2012, focuses on raising the proportion of eigth-grade students leaving middle school at or above a grade-appropriate reading level, and is described as including coaching, inter-site visitation, and workshops for teachers and school leaders, and "other resources" (NYCDOE, 2016c). This, too, introduces a more tightly focused goal and desired outcome for collaboration. 'Showcase Schools' is an initiative in which selected schools host three visits, open to all NYCDOE educators, "where visitors will learn about specific practices, the conditions that contribute to their growth and sustainability, and the story of how they came to be" (NYCDOE, 2016d); it is unclear whether this particular initiative involves clearly bounded goals, and certainly the commitment of visitors is not required beyond a single visit, but the emphasis on celebrating schools' successes might be seen as a way of building buy-in and perceived value with regard to collaboration between schools more generally. Each of these initiatives involves between 37 (for 'Showcase Schools') and 121 (for 'Learning Partners') participating schools in 2016-17; as noted above, a study focusing on one or more of these forms of inter-school collaboration might be a natural follow-up to the present study and an opportunity to further develop the implications and recommendations for policy given above.

While the research presented in this thesis does not ultimately offer evidence that can make strong claims for or against the use of school networks in general, it offers a new contribution to the existing evidence base on school networks and inter-school collaboration, and can hopefully serve as a basis for useful future research as well as evidence-based policy decisions. In particular the results suggests that grouping schools into new organisational structures is not sufficient to foster inter-school collaboration, nor is there reason to expect that such groupings would contribute to school improvement or student achievement. More careful attention to the development of networks, to ensure that such processes ofcollaboration and professional learning are developed and embedded, would be needed to further test the potential role of networks in raising the achievement (or enhancing other important social and affective outcomes) of students.

REFERENCES

Adler, L., Cragin, J., & Searls, P. (1995). *The Los Angeles Area Business/Education Partnership: A Study of the Impact of a Community Based School to Work Program for High Risk Youth.*

Aitkin, M., & Longford, N. (1986). Statistical Modelling Issues in School Effectiveness Studies. *Journal of the Royal Statistical Society. Series A (General)*, *149*(1), 1–43.

Antoniou, P., & Kyriakides, L. (2011). The impact of a dynamic approach to professional development on teacher instruction and student learning: results from an experimental study. *School Effectiveness and School Improvement*, *22*(3), 291–311.

Atkinson, M., Springate, I., Johnson, F., & Halsey, K. (2007). *Inter-school collaboration: A literature review*. Slough: NFER.

Averch, H. A. (1971). *How effective is schooling? A critical review and synthesis of research findings.* Santa Monica, Cal.: Rand Corp.

Ballou, D., Sanders, W., & Wright, P. (2004). Controlling for Student Background in Value-Added Assessment of Teachers. *Journal of Educational and Behavioral Statistics*, *29*(1), 37–65.

Bass, B. M., Drenth, P., & Weissenberg, P. (1987). *Advances in organizational psychology : an international review*. Newbury Park: Sage Publications.

Bassey, M. (1999). *Case study research in educational settings*. Buckingham [England]; Philadelphia: Open University Press.

Battauz, M., Bellio, R., & Gori, E. (2011). Covariate Measurement Error Adjustment for Multilevel Models With Application to Educational Data. *Journal of Educational and Behavioral Statistics*, *36*(3), 283–306.

Bazeley, P. (2012). Integrative Analysis Strategies for Mixed Data Sources. *American Behavioral Scientist*, *56*(6), 814–828.

Bazeley, P., & Jackson, K. (2013). *Qualitative data analysis with NVivo*. London: SAGE.

Bell, M., Jopling, M., Cordingley, P., Firth, A., King, E., & Mitchell, H. (2006). *What is the impact on pupils of networks that include at least three schools? What additional benefits are there for practitioners, organisations and the communities they serve?* Nottingham: National College for School Leadership.

Bielefeldt, T., Moursund, D., Underwood, S., & Underwood, D. (1999). Connected Learning Communities: Findings from the Road Ahead Program, 1995-1997. In *Paper presented at the American Educational Research Association Annual Conference*. Montreal, Quebec, Canada: April 19–23, 1999.

Black-Hawkins, K., McIntyre, D., & McLaughlin, C. (2004). Researching teachers, researching schools, researching networks : a review of the literature. Cambridge: Univ. of Cambridge.

Borman, G., Hewes, G., Overman, L., & Brown, S. (2003). Comprehensive school reform and achievement: A meta-analysis. *Review of Educational Research*, *73*(2), 125–30.

British Educational Research Association. (2011). Ethical guidelines for educational research.

Brooks, R., Te Riele, K., & Maguire, M. (2014). *Ethics and education research*. London: SAGE.

Brophy, J. E., & Good, T. L. (1970). Teachers' communication of differential expectations for children's classroom performance: Some behavioral data. *Journal of Educational Psychology*, *61*(5), 365–374.

Browne, W. J. (2015). MCMC Estimation in MLwiN v2.32. Centre for Multilevel Modeling, University of Bristol.

Browne, W. J., & Draper, D. (2006). A comparison of Bayesian and likelihood-based methods for fitting multilevel models, 473–514.

Browne, W. J., Goldstein, H., & Rasbash, J. (2001). Multiple membership multiple classification (MMMC) models. *Statistical Modelling*, *1*(2), 103–124.

Bryk, A. S., & Raudenbush, S. W. (1988). Toward a more appropriate conceptualisation of research on school effects: A three-level hierarchical linear model. In R. D. Bock (Ed.), *Multilevel analysis of educational data* (pp. 159–204). San Diego; London: Academic Press.

Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models: Applications and data analysis methods*. Newbury Park, Calif.: Sage.

Bryk, A. S., Raudenbush, S. W., & Congdon, R. T. (1986). MM4: Hierarchical linear and nonlinear modeling with the MM/2L and MM/3L programs. Chicago, IL: Scientific Software Internation.

Bryman, A. (2007). Barriers to Integrating Quantitative and Qualitative Research. *Journal of Mixed Methods Research*, *1*(1), 8–22.

Bulkley, K. E. (2010). Introduction -- Portfolio Management Models in Urban School Reform. In K. E. Bulkley, J. R. Henig, & H. M. Levin (Eds.), *Between public and private: Politics, government, and the new portfolio models for urban school reform* (pp. 3–26). Cambridge, Mass.: Harvard Education Press.

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

Chapman, C. (2008). Towards a framework for school-to-school networking in challenging circumstances. *Educational Research*, *50*(4), 403–420.

Chapman, C. (2013). From one school to many. *Educational Management Administration & Leadership*, *43*(1), 46–60.

Chapman, C., & Harris, A. (2004). Improving schools in difficult contexts: Towards a differentiated approach. *British Journal of Educational Studies*, *52*(4), 417–431.

Chapman, C., & Muijs, D. (2013). Collaborative School Turnaround: A Study of the Impact of School Federations on Student Outcomes. *Leadership and Policy in Schools*, *12*(3), 200–226.

Chapman, C., & Muijs, D. (2014). Does school-to-school collaboration promote school improvement? A study of the impact of school federations on student outcomes. *School Effectiveness and School Improvement*, *25*(3), 351–393.

Clark, T. A., & McCarthy, D. P. (1983). School Improvement in New York City: The Evolution of a Project. *Educational Researcher*, *12*(4), 17–24.

Coe, R., & Fitz-Gibbon, C. T. (1998). School Effectiveness Research: criticisms and recommendations. *Oxford Review of Education*, *24*(4), 421–438.

Coffey, A., & Atkinson, P. (1996). *Making sense of qualitative data : complementary research strategies*. Thousand Oaks, CA: Sage.

Coleman, A. (2012). The significance of trust in school-based collaborative leadership. *International Journal of Leadership in Education*, *15*(1), 79–106.

Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, J., Mood, A. M., Weinfeld, F. D., & York, R. L. (1966). *Equality of educational opportunity*. Washington, D.C.: U.S. Dept. of Health, Education, and Welfare, Office of Education.

Common Core State Standards Initiative. (2015). Development process. Retrieved January 13, 2016, from http://www.corestandards.org/about-the-standards/development-process/

Common Core State Standards Initiative. (2016). About the Standards. Retrieved September 3, 2016, from http://www.corestandards.org/about-the-standards/

Corcoran, S., & Pai, G. (2013). *Unlocking New York City's high school Progress Report*. New York: Institute for Education and Social Policy, New York University.

Creemers, B. (1994). *The effective classroom*. London: Cassell.

Creemers, B., & Kyriakides, L. (2008). *The dynamics of educational effectiveness : a contribution to policy, practice and theory in contemporary schools*. London; New York: Routledge.

Creemers, B., & Kyriakides, L. (2010). *Improving quality in education : dynamic approaches to school improvement*. London: Routledge.

Creemers, B., & Kyriakides, L. (2012). *Improving quality in education : dynamic approaches to school improvement*. Abingdon, Oxon; New York: Routledge.

Creemers, B., Kyriakides, L., & Antoniou, P. (2013). A dynamic approach to school improvement: main features and impact. *School Leadership & Management*, *33*(2), 114–132.

Creemers, B., Kyriakides, L., & Sammons, P. (2010). *Methodological Advances in Educational Effectiveness Research*. Milton Park, Abingdon, Oxon; New York: Routledge.

Creemers, B., & Reezigt, G. J. (1992). *Evaluation of educational effectiveness*. Groningen: ICO, Interuniversitair Centrum voor Onderwijsonderzoek.

Creswell, J. W. (2008). *Educational research : planning, conducting, and evaluating*

quantitative and qualitative research (3rd ed.). Upper Saddle River, N.J.: Pearson/Merrill Prentice Hall.

Creswell, J. W., & Plano Clark, V. L. (2011). *Designing and conducting mixed methods research* (2nd ed.). Los Angeles: SAGE Publications.

Creswell, J. W., Plano Clark, V. L., Gutmann, M. L., & Hanson, W. E. (2003). Advanced mixed methods research designs. In A. Tashakkori & C. Teddlie (Eds.), *Handbook of mixed methods in social and behavioral research* (pp. 209–240). Thousand Oaks, CA: Sage Publications.

Creswell, J. W., Plano Clark, V. L., Gutmann, M. L., & Hanson, W. E. (2008). Advanced mixed methods research designs. In V. L. Plano Clark & J. W. Creswell (Eds.), *The mixed methods reader* (pp. 161–196). Thousand Oaks, Calif.: Sage Publications.

Crosser, S. L. (1991). Summer Birth Date Children: Kindergarten Entrance Age and Academic Achievement. *The Journal of Educational Research*, *84*(3), 140–146.

CTB/McGraw Hill. (2010a). New York State testing program 2010: English Language Arts, Grades 3-8, technical report. Monterey, CA: CTB/McGraw Hill.

CTB/McGraw Hill. (2010b). New York State testing program 2010: Mathematics, Grades 3-8, technical report. Monterey, CA: CTB/McGraw Hill.

CTB/McGraw Hill. (2011a). New York State testing program 2011: English Language Arts, Grades 3-8, technical report. Monterey, CA: CTB/McGraw Hill.

CTB/McGraw Hill. (2011b). New York State testing program 2011: Mathematics, Grades 3-8, technical report. Monterey, CA: CTB/McGraw Hill.

Cuban, L. (1984). Transforming the Frog into a Prince: Effective Schools Research, Policy, and Practice at the District Level. *Harvard Educational Review*, *54*(2), 129–151.

D'Haenens, E., Van Damme, J., & Onghena, P. (2010). Multilevel exploratory factor analysis: illustrating its surplus value in educational effectiveness research. *School Effectiveness and School Improvement*, *21*(2), 209–235.

Danielson, C. (2007). *Enhancing professional practice: A framework for teaching* (2nd ed.). Alexandria, Va.: Association for Supervision and Curriculum Development.

Dawsey, J. (2016, March 15). NYC mayor Bill De Blasio cements ties with teachers union. *Wall Street Journal*. New York. Retrieved from http://www.wsj.com/articles/nyc-mayor-bill-de-blasio-cements-ties-with-teachers-union-1458090230

Day, C., Sammons, P., & Gu, Q. (2008). Combining Qualitative and Quantitative Methodologies in Research on Teachers' Lives, Work, and Effectiveness: From Integration to Synergy. *Educational Researcher*, *37*(6), 330–342.

Day, C., Sammons, P., Kington, A., Regan, E., Ko, J., Brown, E., … Robertson, D. (2008). *Effective Classroom Practice: A mixed-methods study of influences and outcomes: Full research report, ESRC End of award report*. Swindon: ESRC.

Day, C., Sammons, P., Leithwood, K., Hopkins, D., Gu, Q., Brown, E., & Ahtaridou.

(2011). *Successful school leadership : linking with learning and achievement*. Maidenhead, England: Open University Press.

de Jong, R., Westerhof, K. J., & Kruiter, J. H. (2004). Empirical Evidence of a Comprehensive Model of School Effectiveness: A Multilevel Study in Mathematics in the 1st Year of Junior General Education in The Netherlands. *School Effectiveness and School Improvement*, *15*(1), 3–31.

Demetriou, D., & Kyriakides, L. (2012). The Impact of School Self-Evaluation upon Student Achievement: A Group Randomisation Study. *Oxford Review of Education*, *38*(2), 149–170.

den Brok, P., Brekelmans, M., & Wubbels, T. (2006). Multilevel Issues in Research Using Students' Perceptions of Learning Environments: The Case of the Questionnaire on Teacher Interaction. *Learning Environments Research*, *9*(3), 199–213.

Department of Health and Human Services. Protection of human subjects (2009). Retrieved from http://www.hhs.gov/ohrp/policy/ohrpregulations.pdf

Dimmock, C. (1995). Restructuring for School Effectiveness: Leading, Organising and Teaching for Effective Learning. *Educational Management & Administration*, *23*(1), 5–18.

Drechsler, J. (2015). Multiple Imputation of Multilevel Missing Data—Rigor Versus Simplicity. *Journal of Educational and Behavioral Statistics*, *40*(1), 69–95.

Editorial Board. (2014, February 14). The Common Core in New York. *New York Timesk*. New York. Retrieved from http://www.nytimes.com/2014/02/15/opinion/the-common-core-in-new-york.html

Edmonds, R. R. (1979). Effective schools for the urban poor. *Educational Leadership*, *37*(1), 15–27.

Enders, C. K. (2010). *Applied missing data analysis*. New York: Guilford Press.

Fitz-Gibbon, C. T., Tymms, P. B., & Hazlewood, R. D. (1989). Performance indicators and information systems. In *School effectiveness and improvement: Selected proceedings of the First International Congress for School Effectiveness* (pp. 141–52). Groningen, Netherlands: RION.

Freedberg, L. (2016). New York stumbles, California advances on Common Core implementation. Retrieved September 3, 2016, from https://edsource.org/2016/new-york-stumbles-california-advances-on-common-core-implementation/92986

Fruchter, N., Ardvisson, T. S., Mokhtar, C., Beam, J., & Simmons, W. (2015). *Demographics and performance in New York City's school networks: An initial inquiry*. Providence, RI.

Fullan, M. (1992). *Successful school improvement : the implementation perspective and beyond*. Buckingham [England]; Philadelphia: Open University Press.

Fullan, M. (2006). *Turnaround leadership*. San Francisco, CA: Jossey-Bass.

Gelman, A., & Shalizi, C. R. (2013). Philosophy and the practice of Bayesian statistics.

British Journal of Mathematical and Statistical Psychology, *66*(1), 8–38.

Gettinger, M., Stoiber, K. C., & Lange, J. (1999). Collaborative Investigation of Inclusive Early Education Practices: A Blueprint for Teacher-Researcher Partnership. *Journal of Early Intervention*, *22*(3), 257–265.

Glazer, J. L., & Peurach, D. J. (2013). School Improvement Networks as a Strategy for Large-Scale Education Reform: The Role of Educational Environments. *Educational Policy*, *27*(4), 676–710.

Goldstein, H. (1997). Methods in School Effectiveness Research. *School Effectiveness and School Improvement*, *8*(4), 369–395.

Goldstein, H. (2010). Multilevel Statistical Models. Hoboken, N.J.: John Wiley & Sons.

Goldstein, H. (2011). *Multilevel statistical models*. Hoboken, N.J.: John Wiley & Sons.

Goldstein, H., & Sammons, P. (1997). The influence of secondary and junior schools on sixteen year examination performance : a cross-classified multilevel analysis. *School Effectiveness and School Improvement*, *8*(2), 219–230.

Gorard, S. (2010). Serious doubts about school effectiveness. *British Educational Research Journal*, *36*(5), 745–766.

Graham, J. W., Olchowski, A. E., & Gilreath, T. D. (2007). How Many Imputations are Really Needed? Some Practical Clarifications of Multiple Imputation Theory. *Prevention Science*, *8*(3), 206–213.

Gray, J., Jesson, D., Goldstein, H., Hedger, K., & Rasbash, J. (1995). A Multi-Level Analysis of School Improvement: Changes in Schools' Performance over Time. *Sch. Effectiveness & Sch. Improvement School Effectiveness and School Improvement*, *6*(2), 97–114.

Greenberg, K. H. (1996). *The Cognitive Enrichment Network Education Model (COGNET).*

Greene, J. C. (2007). *Mixed methods in social inquiry*. San Francisco, Calif.: Jossey-Bass.

Greene, J. C., & Caracelli, V. J. (1997). *Advances in mixed-method evaluation : the challenges and benefits of integrating diverse paradigms*. San Francisco: Jossey-Bass Publishers.

Greene, J. C., Caracelli, V. J., & Graham, W. F. (1989). Toward a Conceptual Framework for Mixed-Method Evaluation Designs. *Educational Evaluation and Policy Analysis*, *11*(3), 255–274.

Gronn, P. (2002). Distributed leadership as a unit of analysis. *The Leadership Quarterly*, *13*(4), 423–451.

Gyurko, J., & Henig, J. R. (2010). Strong vision, learning by doing, or the politics of muddling through? In K. E. Bulkley, J. R. Henig, & H. M. Levin (Eds.), *Politics, government, and the new portfolio models for urban school reform* (pp. 91–126). Cambridge, Mass.: Harvard Education Press.

Hadfield, M. (2007). Co-Leaders and Middle Leaders: The Dynamic between Leaders and Followers in Networks of Schools. *School Leadership & Management*, *27*(3), 259–283.

Hallinger, P. (2003). Leading Educational Change: reflections on the practice of instructional and transformational leadership. *Cambridge Journal of Education*, *33*(3), 329–352.

Hallinger, P., & Murphy, J. (1986). The social context of effective schools. *American Journal of Education*, *94*, 328–55.

Harris, A. (2006). From school leader to educational leader. *School Leadership & Management*, *26*(5), 415–417.

Hauser, R. M. (1971). *Socioeconomic background and educational performance.* [Washington: American Sociological Association.

Heck, R., & Hallinger, P. (2010). Testing a longitudinal model of distributed leadership effects on school improvement. *The Leadership Quarterly*, *21*(5), 867–885.

Hedges, L. V, & Olkin, I. (1985). Statistical methods for meta-analysis. Orlando: Academic Press.

Hemphill, C., & Nauer, K. (2010). *Managing by the Numbers: Empowerment and Accountability in New York City's Schools*. New York: Center for New York City Affairs, Milano the New School for Management and Urban Policy.

Hernandez, J. C., & Gebeloff, R. (2013, August 7). Test Scores Sink as New York Adopts Tougher Benchmarks. *New York Times*. New York. Retrieved from http://www.nytimes.com/2013/08/08/nyregion/under-new-standards-students-see-sharp-decline-in-test-scores.html

Hernandez, J. C., & Otterman, S. (2010, November 23). Education chief raises doubts on pick by Bloomberg. *New York Times*. New York. Retrieved from http://www.nytimes.com/2010/11/24/nyregion/24waiver.html

Hesse-Biber, S. N. (2010). Mixed methods research merging theory with practice. New York: Guilford Press.

Higham, R., Hopkins, D., & Matthews, P. (2009). *System leadership in practice*. Maidenhead, England: McGraw Hill/Open University Press.

Hill, P. T., Pierce, L. C., & Guthrie, J. W. (1997). *Reinventing public education : how contracting can transform America's schools*. Chicago: University of Chicago Press.

Hill, P. W., & Rowe, K. J. (1998). Modelling Student Progress in Studies of Educational Effectiveness. *School Effectiveness and School Improvement*, *9*(3), 310–333.

Hopkins, D. (1987). Improving the quality of schooling : lessons from the OECD International School Improvement Project. London; New York: Falmer Press.

Hopkins, D., & Higham, R. (2007). System leadership: mapping the landscape. *School Leadership & Management*, *27*(2), 147–166.

Howe, K. R., & Moses, M. S. (1999). Ethics in Educational Research. *Review of Research in Education*, *24*, 21–59.

Howley-Rowe, C. (2000). *Tinder Elementary a case study of the Quest network*. Charleston, W. Va.; Washington, DC: Office of Educational Research and Improvement / AEL Inc.

Hox, J. J. (2010). *Multilevel analysis : techniques and applications*. New York: Routledge.

Hoy, W. K., & Miskel, C. G. (2013). *Educational administration : theory, research, and practice*. New York; London: McGraw-Hill Higher Education.

Hunter, R. C., & Brown, Frank, Brown, F. (2002). *Challenges of urban education and efficacy of school reform*. Amsterdam; London: JAI.

Hutchings, M., Greenwood, C., Hollingworth, S., Mansaray, A., Rose, A., & Glass, K. (2012). *Evaluation of the City Challenge programme (Research Report DFE-RR215)*. London: Department for Education.

Iatarola, P., Schwartz, A. E., Stiefel, L., & Chellman, C. C. (2008). Small Schools, Large Districts: Small-School Reform and New York City's Students. *Teachers College Record*, *110*(9), 1837–1878.

IBM Corp. (2012). IBM SPSS Statistics for Windows. Armonk, NY: IBM Corp.

Jackson, D. (2006). *Networked learning communities : setting school-to-school collabortion within a system context*. Jolimont, Vic.: Centre for Strategic Education.

Jamieson, I., & Wikeley, F. (2000). Is Consistency a Necessary Characteristic for Effective Schools? *School Effectiveness and School Improvement*, *11*(4), 435–452.

Jencks, C., Smith, M., Acland, H., Bane, M. J., Cohen, D., Gintis, H., … Michelson, S. (1972). *Inequality: A reassessment of the effect of family and schooling in America*. New York: Basic Books.

Jessen, S. B. (2013). Special Education & School Choice: The Complex Effects of Small Schools, School Choice and Public High School Policy in New York City. *Educational Policy*, *27*(3), 427–466.

Johnson, R. B., Onwuegbuzie, A. J., & Turner, L. A. (2007). Toward a Definition of Mixed Methods Research. *Journal of Mixed Methods Research*, *1*(2), 112–133.

Jolliffe, W., & Hutchinson, H. (2007). Implementing Cooperative Learning in a Networked Learning Community. *Education 3-13*, *35*(1), 5–16.

Kahne, J., O'Brien, J., Brown, A., & Quinn, T. (2001). Leveraging Social Capital and School Improvement: The Case of a School Network and a Comprehensive Community Initiative in Chicago. *Educational Administration Quarterly*, *37*(4), 429–461.

Katz, S., & Earl, L. (2010). Learning about networked learning communities. *School Effectiveness and School Improvement*, *21*(1), 27–51.

Kelleher, M. (2014a). *New York City's Children First: Lessons in school reform*.

Washington, D.C.: Center for American Progress.

Kelleher, M. (2014b). *The rise of networks: How decentralised management is improving schools*. Washington, D.C.: Center for American Progress.

Kelly, A. (2012). Measuring “Equity”; and “Equitability”; in School Effectiveness Research. *British Educational Research Journal*, *38*(6), 977–1002.

Kington, A., Sammons, P., Day, C., & Regan, E. (2011). Stories and Statistics: Describing a Mixed Methods Study of Effective Classroom Practice. *Journal of Mixed Methods Research*.

Kyriakides, L. (2002). A Research-Based Model for the Development of Policy on Baseline Assessment. *British Educational Research Journal*, *28*(6), 805–826.

Kyriakides, L. (2004). Differential School Effectiveness in Relation to Sex and Social Class: Some Implications for Policy Evaluation. *Educational Research and Evaluation*, *10*(2), 141–161.

Kyriakides, L. (2005). Extending the Comprehensive Model of Educational Effectiveness by an Empirical Investigation. *School Effectiveness and School Improvement*, *16*(2), 103–152.

Kyriakides, L., & Creemers, B. (2008). Using a multidimensional approach to measure the impact of classroom-level factors upon student achievement: a study testing the validity of the dynamic model. *School Effectiveness and School Improvement*, *19*(2), 183–205.

Kyriakides, L., & Creemers, B. (2011). Can Schools Achieve Both Quality and Equity? Investigating the Two Dimensions of Educational Effectiveness. *Journal of Education for Students Placed at Risk*, *16*(4), 237–254.

Laferriere, T., Montane, M., Gros, B., Alvarez, I., Bernaus, M., Breuleux, A., … Lamon, M. (2010). Partnerships for Knowledge Building: An Emerging Model. *Canadian Journal of Learning and Technology*, *36*(1).

Layton, L., & Chandler, M. A. (2013, November 8). De Blasio aims to reverse education policies in New York. *Washington Post*. New York. Retrieved from https://www.washingtonpost.com/local/education/de-blasio-aims-to-reverse-education-policies-in-new-york/2013/11/08/5de6d24c-47f2-11e3-b6f8-3782ff6cb769_story.html

Lee, V. E., & Bryk, A. S. (1989). A Multilevel Model of the Social Distribution of High School Achievement. *Sociology of Education*, *62*(3), 172–192.

Leech, N. L., & Onwuegbuzie, A. J. (2009). A typology of mixed methods research designs. *Quality & Quantity*, *43*(2), 265–275.

Leithwood, K., Harris, A., & Hopkins, D. (2008). Seven strong claims about successful school leadership. *School Leadership and Management*, *28*(1), 27–42.

Leithwood, K., & Jantzi, D. (2000). The effects of transformational leadership on organizational conditions and student engagement with school. *Journal of Educational Administration*, *38*(2).

Leithwood, K., & Jantzi, D. (2006). Transformational School Leadership for Large-Scale Reform: Effects on Students, Teachers, and Their Classroom Practices. *School Effectiveness and School Improvement*, *17*(2), 201–227.

Leithwood, K., & Louis, K. S. (1998). *Organizational learning in schools*. Lisse, Netherlands; Exton, PA: Swets & Zeitlinger Publishers.

Leithwood, K., Louis, K. S., Anderson, S., & Wahlstrom, K. (2004). *How Leadership Influences Student Learning. Review of Research*. Minneapolis, MN: The Wallace Foundation.

Leithwood, K., Louis, K. S., Wahlstrom, K. L., Anderson, S. E., Michlin, M., Mascall, B., … Moore, S. (2009). *Learning from Leadership: Investigating the Links to Improved Student Learning (Final Report of Research to the Wallace Foundation)*. University of Minnesota.

Leithwood, K., & Montgomery, D. J. (1986). *Improving principal effectiveness : the principal profile*. Toronto, Ont.: Ontario Institute for Studies in Education.

Lenkeit, J. (2012). Effectiveness measures for cross-sectional studies: a comparison of value-added models and contextualised attainment models. *School Effectiveness and School Improvement*, *24*(1), 1–25.

Lenkeit, J., & Caro, D. H. (2014). Performance status and change – measuring education system effectiveness with data from PISA 2000–2009. *Educational Research and Evaluation*, *20*(2), 146–174.

Lin, N. (1999). Building a network theory of social capital. *Connections*, *22*(1), 28–51.

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Beverly Hills, Calif.: Sage Publications.

Lindorff-Vijayendran, A. (2014). School support networks in New York City's portfolio school district: A mixed-methods study. Mixed Methods International Research Association.

Lindorff-Vijayendran, A. (2015a). Exploring variation in student attainment in a network structure for school support: A multilevel analysis. Cincinnati, OH: International Congress for School Effectiveness and School Improvement.

Lindorff-Vijayendran, A. (2015b). Networks for school support and improvement: A mixed methods study. Limassol, Cyprus: European Association for Research in Learning and Instruction.

Lindorff-Vijayendran, A. (2015c). Principals' experiences and perceptions of school support networks in New York City. Chicago, IL: American Educational Research Association.

Lindsay, G., Muijs, D., Harris, A., Chapman, C., Arweck, E., & Goodall, J. (2007). *Final Report of the Evaluation of the Federations Policy*. London: Department for Children, Schools and Families: Department for Children, Schools and Families.

Little, R. J. A. (1988). A Test of Missing Completely at Random for Multivariate Data with Missing Values. *Journal of the American Statistical Association*, *83*(404), 1198–

1202.

Little, R. J. A., & Rubin, D. B. (2002). *Statistical analysis with missing data*. Hoboken, N.J.: Wiley.

Longford, N. T. (1986). VARCL: Interactive software for variance components analysis. *The Professional Statistician*, *5*, 28–32.

Louis, K. S., Dretzke, B., & Wahlstrom, K. (2010). How does leadership affect student achievement? Results from a national US survey. *School Effectiveness and School Improvement*, *21*(3), 315–336.

Luyten, H., & Sammons, P. (2010). Multilevel modeling. In B. P. M. Creemers, L. Kyriakides, & P. Sammons (Eds.), *Methodological advances in educational effectiveness research*. New York: Routledge.

MacEachern, S. N., & Berliner, L. M. (1994). Subsampling the Gibbs Sampler. *The American Statistician*, *48*(3), 188–190.

Madaus, G. G., Kellagham, T., Rakow, E. A., & King, D. (1979). The sensitivity of measures of school effectiveness. *Harvard Educational Review*, *4*, 207–30.

Marks, H. M., & Louis, K. S. (1997). Does Teacher Empowerment Affect the Classroom? The Implications of Teacher Empowerment for Instructional Practice and Student Academic Performance. *Educational Evaluation and Policy Analysis*, *19*(3), 245–275.

Marsh, H., Ludtke, O., Robitzsch, A., Trautwein, U., Asparouhov, T., Muthen, B., & Nagengast, B. (2009). Doubly-Latent Models of School Contextual Effects: Integrating Multilevel and Structural Equation Approaches to Control Measurement and Sampling Error. *Multivariate Behavioral Research*, *44*(6), 764–802.

Maxcy, S. J. (2003). Pragmatic threads in mixed methods research in the social sciences: The search for multiple modes of inquiry and the end of the philosophy of formalism. In A. Tashakkori & C. Teddlie (Eds.), *Handbook of mixed methods in social and behavioral research* (pp. 51–90). Thousand Oaks, CA: Sage Publications.

May, T. (2011). Positioning and Belonging. In T. May & B. Perry (Eds.), *Social Research & Reflexivity: Content, Consequences and Context* (pp. 101–123). Sage Publications.

McCormack-Larkin, M., & Kritek, W. J. (1982). Milwaukee's Project RISE. *Educational Leadership*, *40*(3), 16–21.

McPhillips, M., & Jordan-Black, J.-A. (2009). The effect of month of birth on the attainments of primary and secondary school pupils. *British Journal of Educational Psychology*, *79*(3), 419–438.

Melhuish, E., Sylva, K., Sammons, P., Siraj-Blatchford, I., Taggart, B., Phan, M., & Malin, A. (2008). Preschool Influences on Mathematics Achievement. *Science*, *321*(5893), 1161–1162.

Merriam, S. B. (1988). *Case study research in education : a qualitative approach*. San Francisco: Jossey-Bass.

Montgomery, D. (2001). Increasing Native American Indian Involvement in Gifted Programs in Rural Schools. *Psychology in the Schools*, *38*(5), 467–475.

Morgan, D. L. (2007). Paradigms Lost and Pragmatism Regained: Methodological Implications of Combining Qualitative and Quantitative Methods . *Journal of Mixed Methods Research* , *1*(1), 48–76.

Morse, J. (1991). Approaches to qualitative-quantitative methodological triangulation. *Nursing Research*, *40*, 120–123.

Mortimore, P., Sammons, P., Stoll, L., Lewis, D., & Ecob, R. (1988). *School matters: The junior years*. Somerset: Open Books.

Mosselson, J. (2010). Subjectivity and reflexivity: locating the self in research on dislocation. *International Journal of Qualitative Studies in Education*, *23*(4), 479–494.

Mruck, K., & Breuer, F. (2003). Subjectivity and Reflexivity in Qualitative Research. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research; Vol 4, No 2 (2003): Subjectivity and Reflexivity in Qualitative Research II*. Retrieved from http://www.qualitative-research.net/index.php/fqs/article/view/696/1504

Muijs, D., Ainscow, M., Chapman, C., & West, M. (2011). Collaboration and Networking in Education. Dordrecht: Springer Science+Business Media B.V.

Muijs, D., Kelly, T., Sammons, P., Reynolds, D., & Chapman, C. (2011). The value of educational effectiveness research: A response to recent criticism. *Research Intelligence*, *114*, 24–5.

Muijs, D., West, M., & Ainscow, M. (2010). Why network? Theoretical perspectives on networking. *School Effectiveness and School Improvement*, *21*(1), 5–26.

Mulford, B., & Silins, H. (2003). Leadership for Organisational Learning and Improved Student Outcomes—What Do We Know? *Cambridge Journal of Education*, *33*(2), 175–195.

Mulford, B., & Silins, H. (2011). Revised models and conceptualisation of successful school principalship for improved student outcomes. *International Journal of Educational Management*, *25*(1), 61–82.

Mulgrew, M. (2012). Testimony on the Department of Education's Children First Networks. Speech, New York: United Federation of Teachers. Retrieved from http://www.uft.org/testimony/testimony-department-educations-children-first-networks

Munch, R. (2001). History of functionalism. In N. Smelser & B. Baltes (Eds.), *International encyclopedia of the social and behavioral sciences* (pp. 5838–5844). Dordrecht: Elsevier.

Murphy, J. (1992). School effectiveness and school restructuring: contributions to educational improvement. *School Effectiveness and School Improvement*, *3*(2), 90–109.

Muthen, B. (1990). *Mean and covariance structure analysis of hierarchical data*. Los

Angeles: UCLA.

Muthen, B. (1994). Multilevel Covariance Structure Analysis. *Sociological Methods & Research*, *22*(3), 376–398.

Muthen, L. K., & Muthen, B. (2015). MPlus user's guide. Seventh edition. Los Angeles, CA: Muthen & Muthen.

Nadelstern, E. (2012). *The Evolution of School Support Networks in New York City*. Bothell, WA.

Nathanson, L., Corcoran, S., & Baker-Smith, C. (2013). *High school choice in New York City: A report on the school choices and placements of low-achieving students*.

National Governors Association. (2015). College and career ready training standards. Retrieved September 3, 2016, from http://www.nga.org/cms/home/nga-center-for-best-practices/center-divisions/center-issues/page-edu-issues/col2-content/main-content-list/college-and-career-training-read.html

New York City Department of Education. (2009). Educator guide: The New York City Progress Report, Elementary/Middle school.

New York City Department of Education. (2010). Educator guide: The New York City Progress Report, Elementary/middle.

New York City Department of Education. (2013). Our structure for supporting schools.

New York City Department of Education. (2014). Testing accommodations - guidance for students and families. Retrieved from http://schools.nyc.gov/NR/rdonlyres/D0479BAC-2FA2-43C2-BB64-073FCEE7A99C/0/TestingAccommodations_Families_FINAL.pdf

New York State Education Department. (2013). Approved teacher practice rubrics for New York State. Retrieved September 10, 2013, from http://usny.nysed.gov/rttt/teachers-leaders/practicerubrics/home.html#ATPR

Ng, H. L., & Koretz, D. (2015). Sensitivity of School-Performance Ratings to Scaling Decisions. *Applied Measurement in Education*, *28*(4), 330–349.

Nuttall, D. L., Goldstein, H., Prosser, R., & Rasbash, J. (1989). Differential school effectiveness. *International Journal of Educational Research*, *13*(7), 769–76.

NYCDOE. (2008). *Children First: A bold, common-sense plan to create great schools for all New York City children*. New York.

NYCDOE. (2012). 2011-12 Network Performance Ratings. Retrieved from http://schools.nyc.gov/NR/rdonlyres/FA94D9B1-FEB8-40B2-AFF5-4B52384E3262/146228/NetworkPerformanceRatingsforDistributionFINAL.pdf

NYCDOE. (2015). Chancellor Fariña Announces Streamlined School Support System to Lift Struggling Schools and Increase Accountability. Press release, New York: NYCDOE. Retrieved from http://schools.nyc.gov/Offices/mediarelations/NewsandSpeeches/2014-2015/Chancellor+Fari%25C3%25B1a+Announces+Streamlined+School+Support+Sy

stem+to+Lift+Struggling+Schools+and+Increas.htm

NYCDOE. (2016a). Chancellor Fariña and the Leadership Team. Retrieved September 3, 2016, from http://schools.nyc.gov/AboutUs/leadership/leadershipteam/Chancellor+Fari%25C3%25B1a.htm

NYCDOE. (2016b). Learning Partners. Retrieved September 15, 2016, from http://schools.nyc.gov/Academics/InterschoolCollaboration/learningpartners/default.htm

NYCDOE. (2016c). Middle School Quality. Retrieved September 15, 2016, from http://schools.nyc.gov/Academics/InterschoolCollaboration/middleschool/default.htm

NYCDOE. (2016d). Showcase Schools. Retrieved September 15, 2016, from http://schools.nyc.gov/Academics/InterschoolCollaboration/showcaseschools/default.htm

Oakley, A. (1981). Interviewing women: A contradiction in terms. In H. Roberts (Ed.), *Doing feminist research* (pp. 30–61). London: Routledge and Kegan Paul.

OECD. (2000). *Knowledge management in the learning society*. Paris.

Onwuegbuzie, A. J., & Johnson, R. B. (2006). The validity issue in mixed research. *Research in the Schools*, *13*(1), 48–63.

Onwuegbuzie, A. J., Johnson, R. B., & Collins, K. M. T. (2009). Call for mixed analysis: A philosophical framework for combining qualitative and quantitative approaches. *International Journal of Multiple Research Approaches*, *3*(2), 114–139. Retrieved from http://search.proquest.com/docview/219052535?accountid=13042

Opdenakker, M.-C., & Van Damme, J. (2001). Relationship between School Composition and Characteristics of School Process and their Effect on Mathematics Achievement. *British Educational Research Journal*, *27*(4), 407–432.

Ortega, L. (2016). *Educational effectiveness and inequalities in Chile: A multilevel accelerated longitudinal study of primary school children's achievement trajectories*. University of Oxford.

Oshima, T. C., & Domaleski, C. S. (2006). Academic Performance Gap between Summer-Birthday and Fall-Birthday Children in Grades K-8. *The Journal of Educational Research*, *99*(4), 212–217.

Palardy, G. J. (2008). Differential school effects among low, middle, and high social class composition schools: a multiple group, multilevel latent growth curve analysis. *School Effectiveness and School Improvement*, *19*(1), 21–49.

Parlett, M. R., & Hamilton, D. (1972). *Evaluation as illumination : a new approach to the study of innovatory programs*. [Edinburgh]: [University of Edinburgh, Centre for Research in the Educational Sciences].

Patton, M. Q. (1980). *Qualitative evaluation methods*. Beverly Hills: Sage Publications.

Pearson. (2012a). *New York State testing program 2012: English Language Arts, Grades*

3-8, technical report.

Pearson. (2012b). *New York State testing program 2012: Mathematics, Grades 3-8, technical report*.

Pearson. (2013). New York State testing program 2013: English Language Arts, Mathematics, Grades 3-8, technical report. Pearson.

Peters, J. (2002). Expecting too much from School/University Collaboration: Identifying Faulty Assumptions. *Asia Pacific Journal of Teacher Education*, *30*(3), 229–242.

Phillips, A. (2011, April 14). Education commissioner approves waiver for Dennis Walcott. *Chalkbeat*. Retrieved from https://www.chalkbeat.org/posts/ny/2011/04/14/education-commissioner-approves-waiver-for-dennis-walcott/#.V_PKGPR5KHc

Pinon, D., Samii-Shore, K., & Batchelder, M. (2002). *Principles of Learning: 2001–2002 Evaluation Report*. Austin, TX.

Press Office of the City of New York. (2014). Mayor De Blasio, UFT reach preliminary agreement on 9-year contract ushering in key new reforms and savings. Retrieved September 3, 2016, from http://www1.nyc.gov/office-of-the-mayor/news/195-14/mayor-de-blasio-uft-reach-preliminary-agreement-9-year-contract-ushering-key-new-reforms-and#/0

Pring, R. A. (2000). *Philosophy of educational research*. London: Continuum.

QSR International. (2012). NVivo qualitative data analysis software. Retrieved from http://www.qsrinternational.com

Rasbash, J., Charlton, C., Browne, W. J., Healy, M., & Cameron, B. (2014). MLwiN Version 2.31. Centre for Multilevel Modeling, University of Bristol.

Raudenbush, S. W. (1989). The analysis of longitudinal, multilevel data. *International Journal of Educational Research*, *13*(7), 721–740.

Raudenbush, S. W., & Bryk, A. S. (1986). A Hierarchical Model for Studying School Effects. *Sociology of Education*, *59*(1), 1–17.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models : applications and data analysis methods*. Thousand Oaks: Sage Publications.

Reezigt, G. J. (2001). *A framework for effective school improvement : final report of the ESI project*. Groningen: GION.

Reezigt, G. J., & Creemers, B. (2005). A comprehensive framework for effective school improvement. *School Effectiveness and School Improvement*, *16*(4), 407–424.

Reynolds, D. (2004). Beyond school effectiveness and school improvement. In A. Harris & N. Bennett (Eds.), *School effectiveness and school improvement: alternative perspectives* (2nd ed., pp. 26–43). London: Continuum.

Reynolds, D., & Stoll, L. (1996). Merging school effectiveness and school improvement: the knowledge bases. In R. Bollen, B. Creemers, D. Hopkins, N. Lagerweij, D.

Reynolds, & L. Stoll (Eds.), *Making good schools: Linking school effectiveness and school improvement*. London; New York: Routledge.

Riley, K., & Jordan, J. (2004). "It Makes Sense to Me": Reforming Classrooms from the Bottom-Up: A Case Study in Change. *Improving Schools*, *7*(3), 227–242.

Robinson, V. M. J. (2010). From Instructional Leadership to Leadership Capabilities: Empirical Findings and Methodological Challenges. *Leadership and Policy in Schools*, *9*(1), 1–26.

Robinson, V. M. J., Lloyd, C. A., & Rowe, K. J. (2008). The Impact of Leadership on Student Outcomes: An Analysis of the Differential Effects of Leadership Types. *Educational Administration Quarterly*, *44*(5), 635–674.

Rubin, D. (1987). *Multiple imputation for nonresponse in surveys*. New York: Wiley.

Rubin, H. J., & Rubin, I. (2012). *Qualitative interviewing : the art of hearing data*. Thousand Oaks, Calif.: SAGE.

Rutter, M., Maughan, B., Mortimore, P., Ouston, J., & Smith, A. (1979). *Fifteen thousand hours: secondary schools and their effects on children*. Cambridge, Mass.: Harvard University Press.

Salokangas, M., & Chapman, C. (2014). Exploring governance in two chains of academy schools: A comparative case study . *Educational Management Administration & Leadership* , *42*(3), 372–386.

Sammons, P. (1999). *School effectiveness : coming of age in the twenty-first century*. Lisse; Exton, PA: Swets & Zeitlinger Publishers.

Sammons, P. (2007). *School effectiveness and equity: Making connections*. Reading.

Sammons, P. (2010a). Equity and educational effectiveness. In P. Peterson, E. Baker, & B. McGaw (Eds.), *International Encyclopedia of Education* (pp. 51–57). Oxford: Elsevier.

Sammons, P. (2010b). The contribution of mixed methods to recent research on educational effectiveness. In A. Tashakkori & C. Teddlie (Eds.), *Handbook of mixed methods in social and behavioral research* (pp. 697–724). Los Angeles, Calif.: SAGE Publications.

Sammons, P., Davis, S., Day, C., & Gu, Q. (2014). Using mixed methods to investigate school improvement and the role of leadership. *Journal of Educational Administration*, *52*(5), 565–589.

Sammons, P., Day, C., Kington, A., Gu, Q., Stobart, G., & Smees, R. (2007). Exploring variations in teachers' work, lives and their effects on pupils: key findings and implications from a longitudinal mixed- method study. *British Educational Research Journal*, *33*(5), 681–701.

Sammons, P., Gu, Q., Day, C., & Ko, J. (2011). Exploring the Impact of School Leadership on Pupil Outcomes: Results from a Study of Academically Improved and Effective Schools in England. *International Journal of Educational Management*,

25(1), 83–101.

Sammons, P., Mujtaba, T., Earl, L., & Gu, Q. (2007). Participation in network learning community programmes and standards of pupil achievement: does it make a difference? *School Leadership & Management*, *27*(3), 213–238.

Sammons, P., Nuttall, D., & Cuttance, P. (1993). Differential school effectiveness: Results from a re-analysis of the Inner London Education Authority's junior school project data. *British Educational Research Journal*, *19*(4), 381–405.

Sammons, P., Siraj- Blatchford, I., Sylva, K., Melhuish, E., Taggart, B., & Elliot, K. (2005). Investigating the Effects of Pre- school Provision: Using Mixed Methods in the EPPE Research. *International Journal of Social Research Methodology*, *8*(3), 207–224.

Sammons, P., Thomas, S., & Mortimore, P. (1997). *Forging links : effective schools and effective departments*. London: Paul Chapman.

Sanders, M. G. (1999). Schools' Programs and Progress in the National Network of Partnership Schools. *Journal of Educational Research*, *92*(4), 220.

Schafer, J. L. (1997). *Analysis of incomplete multivariate data*. London; New York: Chapman & Hall.

Schafer, J. L. (1999). Multiple imputation: a primer. *Statistical Methods in Medical Research* , *8*(1), 3–15.

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, *7*(2), 147–177.

Schafer, J. L., & Olsen, M. K. (1998). Multiple Imputation for Multivariate Missing-Data Problems: A Data Analyst's Perspective. *Multivariate Behavioral Research*, *33*(4), 545–571.

Scheerens, J. (1992). *Effective schooling : research, theory and practice*. London: Cassell.

Scheerens, J. (2013). The Use of Theory in School Effectiveness Research Revisited. *School Effectiveness and School Improvement*, *24*(1), 1–38.

Scheerens, J., & Bosker, R. J. (1997). *The foundations of educational effectiveness*. Oxford, OX; New York, N.Y.: Pergamon.

Shapiro, E. (2015, January 22). Farina to overhaul school-support system, reversing a Bloomberg legacy. *Politico*. New York. Retrieved from http://www.politico.com/states/new-york/city-hall/story/2015/01/farina-to-overhaul-school-support-system-reversing-a-bloomberg-legacy-019043

Simons, H. (2009). *Case study research in practice*. Los Angeles; London: SAGE.

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis : modeling change and event occurrence*. Oxford; New York: Oxford University Press.

Slavin, R. E., & Madden, N. A. (2001). *One million children: Success for all*. Thousand Oaks, CA: Corwin Press.

Smith, D. J., & Tomlinson, S. (1989). *The school effect: A study of multi-racial comprehensives*. London.

Smylie, M. A., & Wenzel, S. A. (2003). *The Chicago Annenberg Challenge: Successes, failures, and lessons for the future. Final technical report of the Annenberg Research Project.* Chicago, IL.

Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel analysis : an introduction to basic and advanced multilevel modeling*. Los Angeles: Sage.

Sondergeld, T. A., & Koskey, K. (2011). Evaluating the impact of an urban comprehensive school reform: An illustration of the need for mixed methods. *Journal of Studies in Educational Evaluation*, *37*(2), 94–107.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Linde, A. van der. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *64*(4), 583. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=bth&AN=7693005&site=ehost-live

Spillane, J., Halverson, R., & Diamond, J. B. (2004). Towards a Theory of Leadership Practice: A Distributed Perspective. *Journal of Curriculum Studies*, *36*(1), 3–34.

Stake, R. E. (1995). *The art of case study research*. Thousand Oaks: Sage Publications.

Stoll, L. (2015). *Three greats for a self-improving school system - pedagogy, professional development and leadership*.

Strand, S. (1999). Ethnic group, sex and economic disadvantage: Associations with pupils' educational progress from Baseline to the end of Key Stage 1. *British Educational Research Journal*, *32*, 209–225.

Strand, S. (2008). *Minority ethnic pupils in the LSYPE: Extension report on performance in public examinations at age 16 (DCSF Research Report RR029)*. Nottingham, UK.

Strand, S. (2010). Do some schools narrow the gap? Differential school effectiveness by ethnicity, gender, poverty, and prior achievement. *School Effectiveness and School Improvement*, *21*(3), 289–314.

Strand, S. (2011). The Limits of Social Class in Explaining Ethnic Gaps in Educational Attainment. *British Educational Research Journal*, *37*(2), 197–229.

Strand, S. (2016). Do some schools narrow the gap? Differential school effectiveness revisited. *Review of Education*, *4*(2), 107–144.

Strand, S., & Lindsay, G. (2009). Evidence of Ethnic Disproportionality in Special Education in an English Population. *The Journal of Special Education*, *43*(3), 174–190.

Strauss, A. L., & Corbin, J. M. (1998). *Basics of qualitative research : techniques and procedures for developing grounded theory*. Thousand Oaks: Sage Publications.

Tashakkori, A., & Teddlie, C. (1998). *Mixed methodology : combining qualitative and quantitative approaches*. Thousand Oaks, Calif.: Sage.

Tashakkori, A., & Teddlie, C. (2003). The past and future of mixed methods research: From data triangulation to mixed model designs. In A. Tashakkori & C. Teddlie (Eds.), *Handbook of mixed methods in social and behavioral research* (pp. 671–701). Thousand Oaks, CA: Sage Publications.

Teddlie, C., & Reynolds, D. (2000). *The international handbook of school effectiveness research*. London; New York: Falmer Press.

Teddlie, C., & Sammons, P. (2010). Applications of mixed methods to the field of educational effectiveness research. In B. Creemers, L. Kyriakides, & P. Sammons (Eds.), *Methodological advances in educational effectiveness research* (pp. 115–152). Milton Park, Abingdon, Oxon; New York: Routledge.

Teddlie, C., & Stringfield, S. (1985). A differential analysis of effectiveness in middle and lower socioeconomic status schools. *Journal of Classroom Interaction*, *20*, 38–44.

Teddlie, C., & Stringfield, S. (1993). *Schools make a difference : lessons learned from a 10-year study of school effects*. New York: Teachers College Press.

Teddlie, C., & Stringfield, S. (2006). A brief history of school improvement research in the USA. In A. Harris & J. Chrispeels (Eds.), *Improving schools and educational systems: International perspectives* (pp. 23–38). New York; London: Routledge.

Teddlie, C., & Tashakkori, A. (2003). Major issues and controversies in the use of mixed methods in the social and behavioral social sciences. In A. Tashakkori & C. Teddlie (Eds.), *Handbook of mixed methods in social and behavioral research* (pp. 3–50). Thousand Oaks, CA: Sage Publications.

Teddlie, C., & Yu, F. (2007). Mixed Methods Sampling: A Typology With Examples . *Journal of Mixed Methods Research* , *1*(1), 77–100.

Televantou, I., Marsh, H. W., Kyriakides, L., Nagengast, B., Fletcher, J., & Malmberg, L.-E. (2015). Phantom effects in school composition research: consequences of failure to control biases due to measurement error in traditional multilevel models. *School Effectiveness and School Improvement*, *26*(1), 75–101.

The Parthenon Group. (2013). *An Assessment of the New York City Department of Education School Support Structure*. Boston.

The University of the State of New York. Amendment of Subpart 30-2 Guidance on New York State's Annual Professional Performance Review for teachers and principals to implement Education Law §3012-c and the Commissioner's regulations (2014). Retrieved from https://www.engageny.org/sites/default/files/resource/attachments/appr-field-guidance.pdf

Thomas, S. (1995). Considering primary school effectiveness: an analysis of 1992 Key Stage 1 results. *The Curriculum Journal*, *6*(3), 279–295.

Thomas, S., & Mortimore, P. (1996). Comparison of value-added models for secondary-school effectiveness. *Research Papers in Education*, *11*(1), 5–33.

Thomas, S., Sammons, P., Mortimore, P., & Smees, R. (1997). Stability and Consistency in Secondary Schools' Effects on Students' GCSE Outcomes over Three Years∗.

School Effectiveness and School Improvement, *8*(2), 169–197.

Thrupp, M. (2001). Recent School Effectiveness Counter-critiques: Problems and possibilities. *British Educational Research Journal*, *27*(4), 443–457.

Thurlow, M. L. (1999). *District partnership approach to inclusion a qualitative evaluation of impact*. Minneapolis, MN; Washington, DC: Institute on Community Integration (UAP), the College of Education & Human Development, University of Minnesota ; U.S. Dept. of Education, Office of Educational Research and Improvement, Educational Resources Information Center.

Timmermans, A. C., & Thomas, S. M. (2014). The impact of student composition on schools' value-added performance: a comparison of seven empirical studies. *School Effectiveness and School Improvement*, 1–12.

Toland, M. D., & De Ayala, R. J. (2005). A Multilevel Factor Analysis of Students' Evaluations of Teaching. *Educational and Psychological Measurement*, *65*(2), 272–296.

van Buren, S. (2011). Multiple imputation of multilevel data. In J. J. Hox & J. K. Roberts (Eds.), *Handbook of advanced multilevel analysis* (pp. 173–196). Milton Park, UK: Routledge.

van Buren, S. (2015). Fully conditional specification. In G. Molenberghs, G. Fitzmaurice, M. G. Kenward, A. Tsiatis, & G. Verbeke (Eds.), *Handbook of missing data methodology*. Boca Raton, FL: Chapman and Hall/CRC.

van de Vijver, Y. H., & Poortinga, F. J. R. (2002). Structural Equivalence in Multilevel Research. *Journal of Cross-Cultural Psychology*, *33*(2), 141–156.

Vygotsky, L., Vygotsky, S., & John-Steiner, V. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, Mass.: Harvard University Press.

Weber, G. (1971). *Inner city children can be taught to read: Four successful schools*. Washington, D.C.: Council for Basic Education.

Willms, J. D. (1992). *Monitoring school performance : a guide for educators*. Washington, D.C.: Falmer.

Willms, J. D. (2010). School composition and contextual effects on student outcomes. *Teachers College Record*, *112*(4), 1008–1037. Retrieved from http://www.scopus.com/inward/record.url?eid=2-s2.0-77649339925&partnerID=40&md5=b9015a5a45e415f44065da12cfb4e1e8

Willms, J. D., & Raudenbush, S. W. (1989). A Longitudinal Hierarchical Linear Model for Estimating School Effects and Their Stability. *Journal of Educational Measurement*, *26*(3), 209–232.

Winters, M. A. (2012). Measuring the effect of charter schools on public school student achievement in an urban environment: Evidence from New York City. *Economics of Education Review*, *31*(2), 293–301.

Winters, M. A. (2015). *Pushed out? Low performing students and New York City's charter schools*. New York.

Wohlstetter, P., Houston, D. M., & Buck, B. (2015). Networks in New York City: Implementing the Common Core . *Educational Policy* , *29*(1), 85–110.

Wohlstetter, P., Malloy, C. L., Chau, D., & Polhemus, J. L. (2003). Improving Schools through Networks: A New Approach to Urban School Reform. *Educational Policy*, *17*(4), 399–430.

Wohlstetter, P., Smith, J., & Gallagher, A. (2013). New York City's Children First Networks: turning accountability on its head. *Journal of Educational Administration*, *51*(4), 528–549.

Woodhouse, G., Yang, M., Goldstein, H., & Rasbash, J. (1996). Adjusting for Measurement Error in Multilevel Analysis. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, *159*(2), 201–212.

Wrigley, T. (2013). Rethinking school effectiveness and improvement: A question of paradigms. *Discourse*, *34*(1), 31–47.

Yin, R. K. (2006). Mixed methods research: Are the methods genuinely integrated or merely parallel? *Research in the Schools*, *13*(1), 41–47.

Yin, R. K. (2014). *Case study research : design and methods*. Los Angeles, Calif.: Sage Publications.

Yukl, G. A. (2006). *Leadership in organizations*. Upper Saddle River, NJ: Pearson/Prentice Hall.

Zetlin, A. G., MacLeod, E., & Michener, D. (1998). *Professional Development of Teachers of Language Minority Students through University-School Partnership.*

APPENDICES

Appendix 1 Missing data analysis

This appendix presents analysis of missing data patterns conducted to inform decisions about missing data treatment. Because of the opening and closure of schools from year to year, missing data for 2009-10, 2010-11, 2011-12, and 2012-13 were treated separately.

Missing data were analysed in the following stages:

1. Frequency and distribution of missing values by variable, case, and cell in the dataset
2. Patterns of missing data
3. Missing data mechanism(s)

Extent of missingness

The pie charts in Figures 11 through 14 below show the proportion of missing values for each year's dataset by variable, case, and cell entry. These reflect student and school level variables; there were not missing values above the school level in any individual dataset.

Figure 10 Extent of missing data: Variables, cases and values (2009-10)

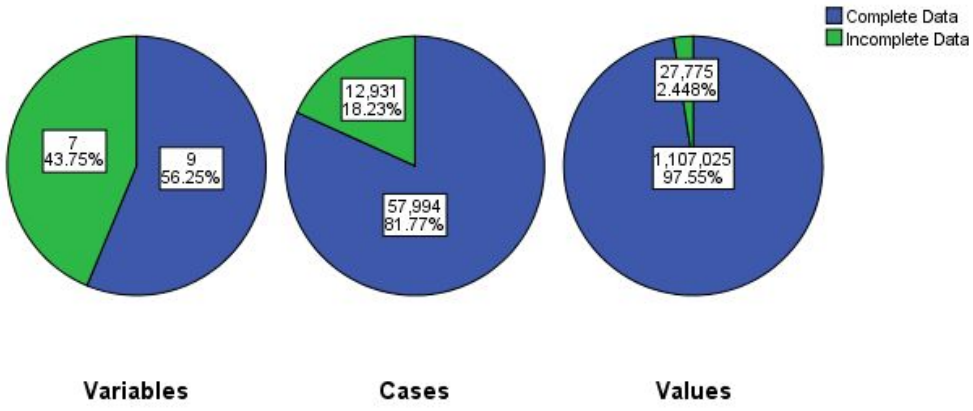


Figure 11 Extent of missing data: Variables, cases and values (2010-11)



Figure 12 Extent of missing data: Variables, cases and values (2011-12)

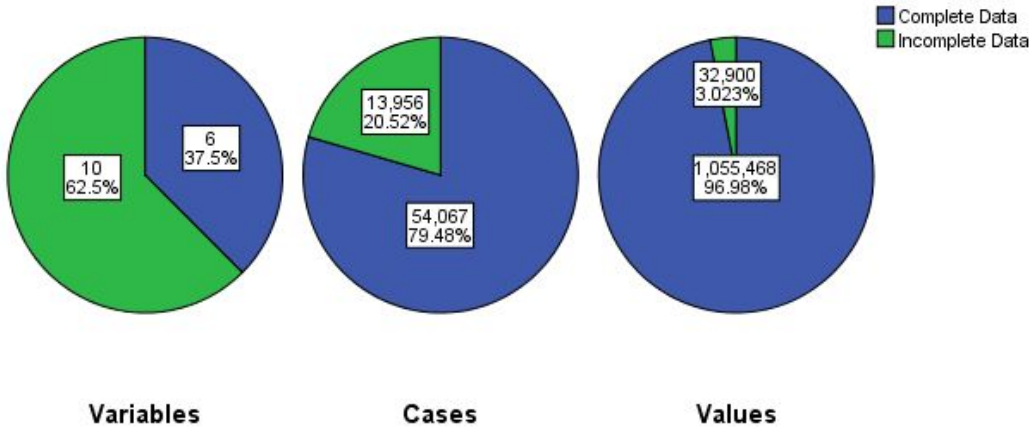
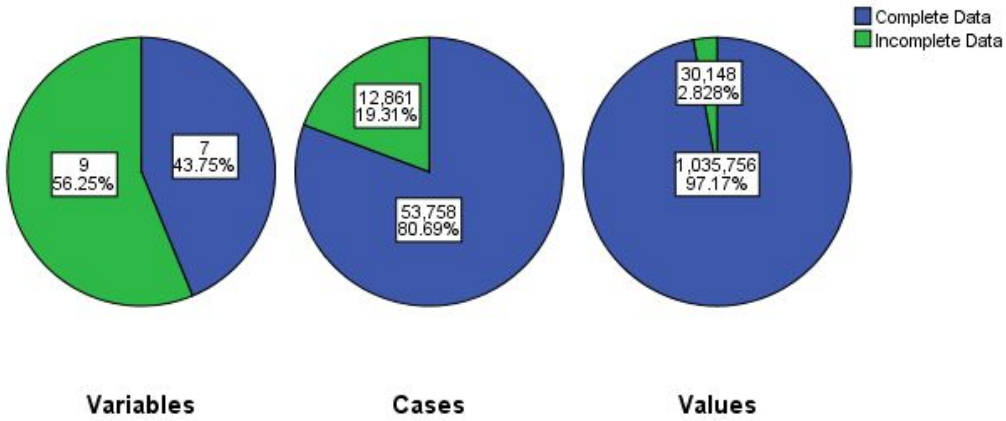


Figure 13 Extent of missing data: Variables, cases and values (2012-13)



As Figures 11 to 14 demonstrate, there were missing values on approximately half of the variables for the dataset for any given year. The proportion of complete cases was for all years approximately 80%, and 2-3% of the cell values in each dataset were blank.

The fact that roughly 20% of cases were missing information suggests that listwise deletion would lead to a substantial reduction in sample size, and the much lower proportion of individual missing values indicates that this would also lead to a sizeable loss of information, as this meant that many of the incomplete cases were only missing one or some of the incomplete variables.

Missing values by variable

Table 52 below summarizes the frequency of missingness for each incomplete variable, along with the corresponding proportion of total cases in a given dataset. By far the most frequently missing variables were 5th grade achievement in literacy and mathematics, followed by 8th grade achievement.

Table 52 Frequency and proportion of missing values by variable and year

	2009-10 Missing		2010-11 Missing		2011-12 Missing		2012-13 Missing	
	N	% of cases	N	% of cases	N	% of cases	N	% of cases
5th grade literacy	11722	16.5%	11825	17.1%	11372	16.7%	10608	15.9%
5th grade math	10588	14.9%	10771	15.6%	10354	15.2%	9354	14.0%
Literacy	3538	5.0%	3224	4.7%	2979	4.4%	3076	4.6%
Math	1712	2.4%	1728	2.5%	1868	2.7%	1935	2.9%
Ethnicity	51	.1%					7	.0%
School mean prior lit.	82	.1%	110	.2%	104	.2%	71	.1%
School mean prior math	82	.1%	110	.2%	104	.2%		
School total enrolment					2039	3.0%	1699	2.6%
Proportion FSM					2039	3.0%	1699	2.6%
Proportion Black/Hispanic					2039	3.0%	1699	2.6%

Blank cells indicate less than 0.1% of the values for a given variable were missing.

Missing data patterns

In addition to understanding how much of the information in a dataset is missing, appropriate decisions regarding missing data treatment also depend on an initial analysis of the patterns by which data is missing, as different approaches may be appropriate for different patterns of missingness. The figures and bar charts below display illustrations of missing data patterns for each year. Each figure displays the patterns of missing data present in a particular dataset and assigns that pattern a number, with red cells highlighting the variables missing in a particular pattern. Each corresponding bar chart shows the percentage of cases to which each numbered pattern in the figure applies, with Pattern 1 always representing a complete-case scenario. For all years, complete cases obviously constitute the largest proportion as noted in the section above (approximately 80%), with the second most common pattern being jointly-missing prior achievement in literacy and mathematics for all years.

Figure 14 Missing data patterns (2009-10)

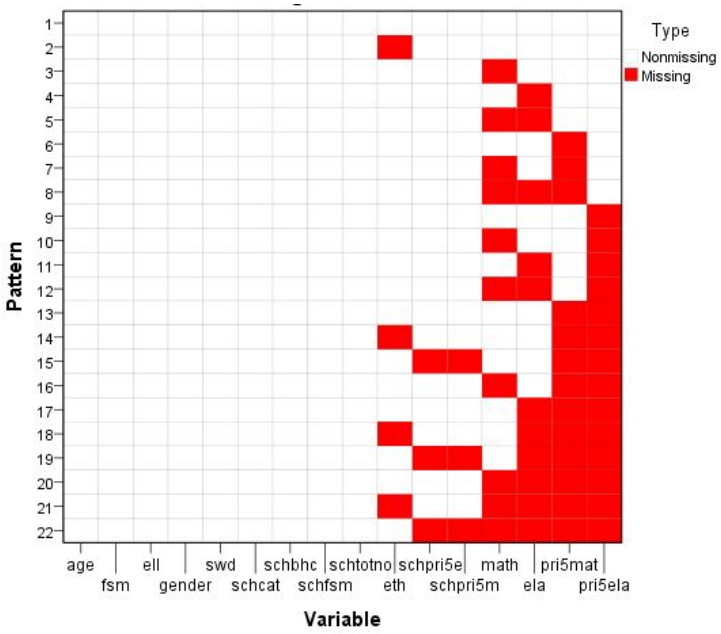


Figure 15 Percentage of cases by pattern (2009-10)

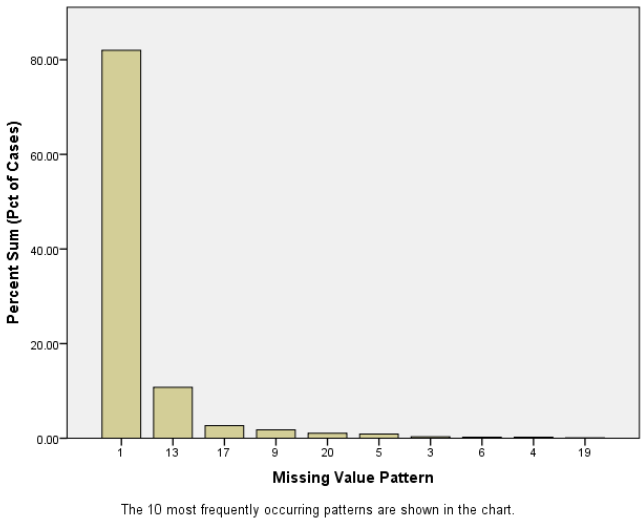


Figure 16 Missing data patterns (2010-11)

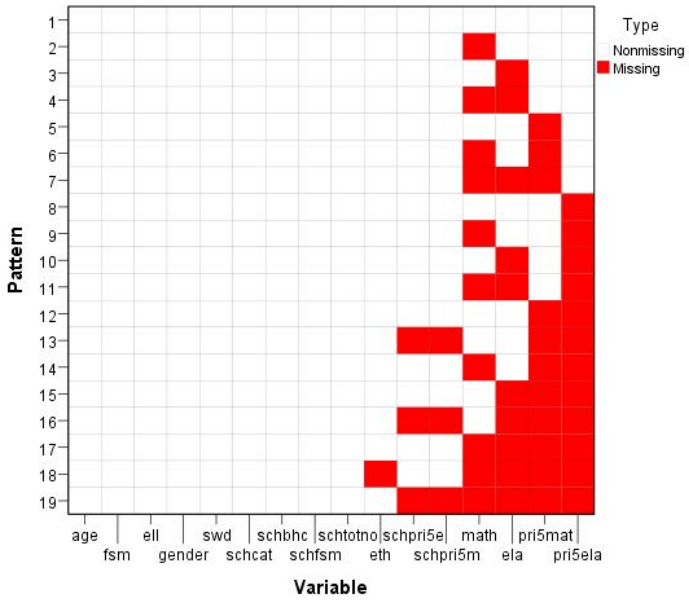


Figure 17 Percentage of cases by pattern (2010-11)

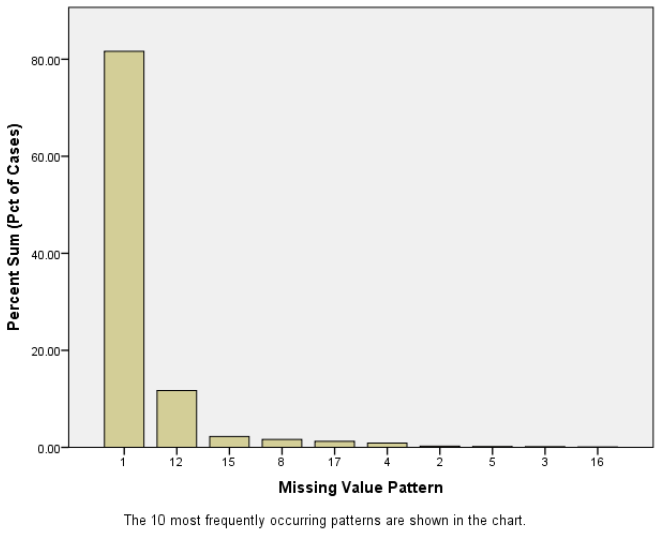


Figure 18 Missing data patterns (2011-12)

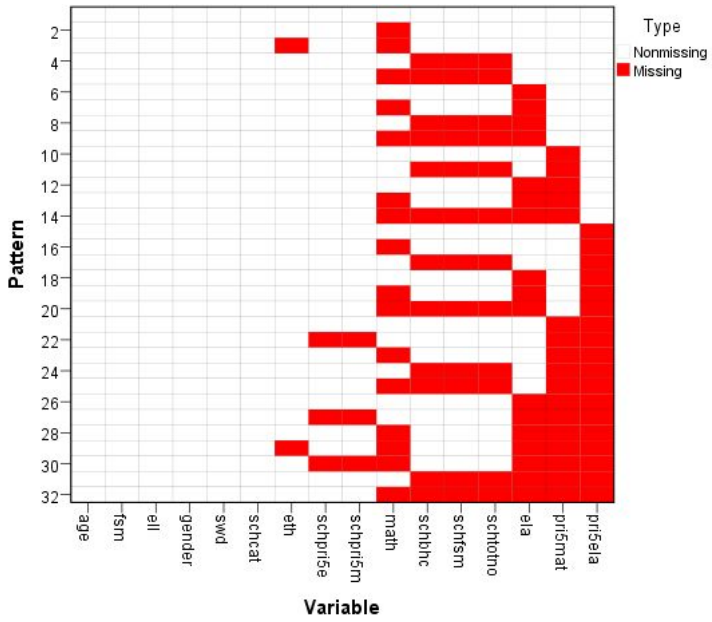


Figure 19 Percentage of cases by pattern (2011-12)

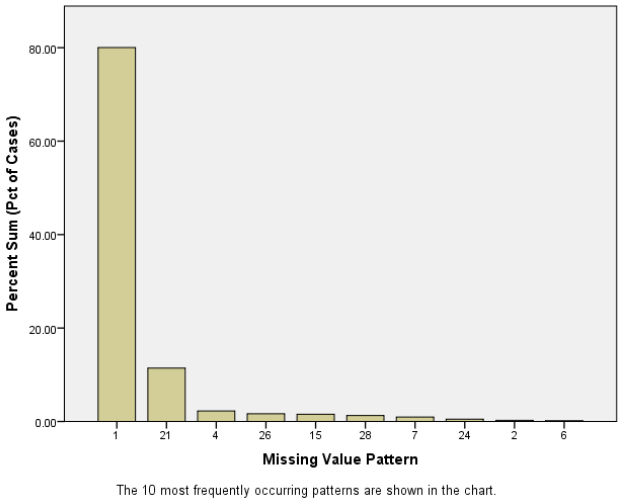


Figure 20 Missing data patterns (2012-13)

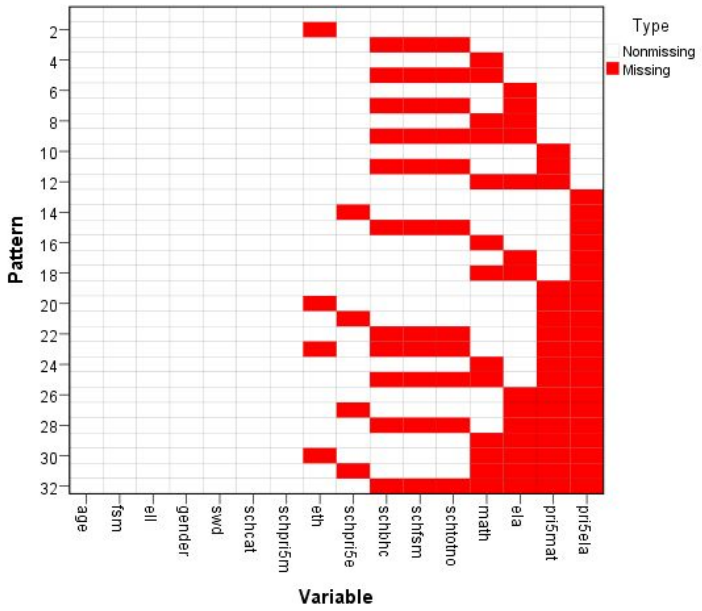
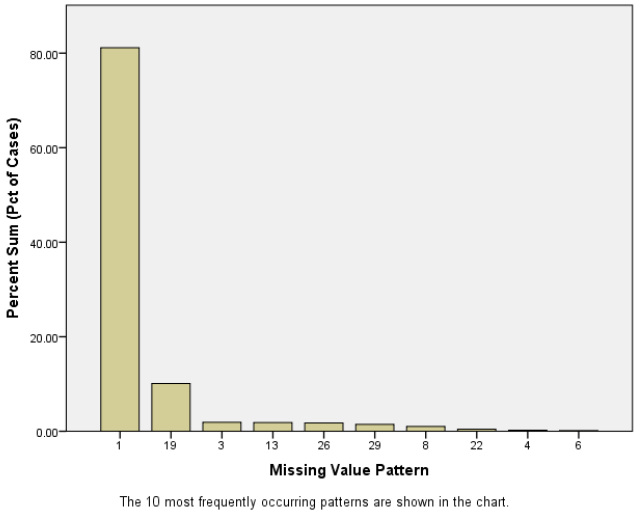


Figure 21 Percentage of cases by pattern (2012-13)



Jointly missing patterns

Tables 53 through 56 provide another way of representing missing data patterns, particularly in the interest of establishing which variables tend to be missing jointly and whether there is evidence of monotonicity of missing variables (i.e. such as one might find in study attrition, where dropout leads to patterns in which a missing measurement at one time point means that measurements at subsequent time points will be missing as well). This appears to be potentially the case in 2009-10, but further ordering and examination of the patterns of missing data suggested that while the below patterns in literacy and math

scores for 5th and 8th grade demonstrated possible monotonicity for that year, the patterns of other missing values on student demographic variables and school variables appeared to be arbitrary, precluding the use of fully conditional specification algorithms for the imputation approach (Joseph L Schafer, 1999). For other years, missing data patterns were similarly not monotonic.

Table 53 Table of missing data patterns by variable (2009-10)

Number of Cases	Missing Patterns				Complete if...
	math	ela	pri5mat	pri5ela	
57994					57994
1268				X	59262
7620			X	X	67044
1878		X	X	X	69085
749	X	X	X	X	70792

Patterns with less than 1% of cases (709 or fewer) are not displayed. "Complete if…" shows the total number of complete cases if cases with a particular pattern of missing values are excluded.

Table 54 Table of missing data patterns by variable (2010-11)

Number of Cases	Missing Patterns				Complete if...
	math	ela	pri5mat	pri5ela	
56386					56386
1138				X	57524
8140			X	X	65777
862	X	X	X	X	69215
1606		X	X	X	67488

Table 55 Table of missing data patterns by variable (2011-12)

Number of Cases	Missing Patterns				Complete if...
	math	ela	pri5mat	pri5ela	
55592					55592
1069				X	56661
8095			X	X	64831
693	X	X			56564
922	X	X	X	X	68021
1204		X	X	X	66155

Table 56 Table of missing data patterns by variable (2012-13)

Number of Cases	Missing Patterns				Complete if...
	math	ela	pri5mat	pri5ela	
55008					55008
1264				X	56272
7022			X	X	63340
706	X	X			55962
1002	X	X	X	X	66612
1237		X	X	X	64678

There were two additional patterns amongst the very small proportion of missing school-level values not included in the above tables. Demographic school composition variables were missing jointly, as these missing values arose from the omission of school records from a single data source (rather than independently missing values within particular schools' records). Missing school demographic variables were present in 2039 cases or 3% of the 2011-12 data set, and 1699 cases or 2.6% of the 2012-13 dataset. Likewise, prior achievement measures at the school level were jointly missing for 82 cases (0.1% of the dataset) in 2009-10 and 110 cases (or 0.2%) in 2010-11 (each of these representing a single school with missing data in the relevant year).

Missing data mechanisms

An overview of missing data mechanisms was given in Chapter 4 (Methods) in discussing the approach used to treat missing data in this study. Little's MCAR test was run separately for student and school variables for each year's data set (Table 57) and demonstrated that the data were not missing completely at random. There is no empirical test for establishing conclusively whether data are missing at random (MAR) or not missing at random (NMAR), so for the purposes of the analyses in this thesis data were assumed to be MAR.

Table 57 Little's MCAR test results by year and variable level

	χ^2	df	p
2009-10 (student)	5436.026	40	<0.001
2009-10 (school)	745.988	3	<0.001
2010-11 (student)	6142.244	40	<0.001
2010-11 (school)	899.935	3	<0.001
2011-12 (student)	6459.783	40	<0.001
2011-12 (school)	3360.943	5	<0.001
2012-13 (student)	7452.580	37	<0.001
2012-13 (school)	3717.104	6	<0.001

Department of Education
Dennis M. Walcott, Chancellor

Institutional Review Board
52 Chambers Street, Room 309
New York, NY 10007
Phone: (212) 374-7659
Fax: (212) 374-5908

APPROVAL TO CONDUCT RESEARCH IN SCHOOLS

To the Principal:

The research study described in the Proposal Submission Form has been approved by the Institutional Review Board (IRB) of the New York City Department of Education. (See the signed Approval Letter) This researcher is now seeking principals willing to cooperate in the study. Please sign below if you agree to have your school participate in this study.

In order to begin the study, the researcher must return this form to IRB@schools.nyc.gov – signed by the principal of each school that will be participating in the study to the IRB, Research and Policy Support Group (RPSG) before data collection begins.

NOTE:

Researchers who need to be in schools must have fingerprints on file at the Department of Education prior to field work.
Where data collection includes information from DOE administrative records, a data request must be submitted to RPSGresearch@schools.nyc.gov. Researchers may not request school or individual student records from school personnel.

Researcher/Principal Investigator: Ariel Lindorff-Vijayendran
Title of Study: Support organizations and school improvement: A mixed-methods study of roles and effects of school support organizations and networks in the context of New York City public schools (file #452)

Research Will Involve:

Cooperating School	School ATS Code (DBN district-boro-school number)	Grade (s)	Number of Classes	Number of Staff/Pupils	Start Date of Data Collection

_____ _____ _____
Principal's Signature *Date* *School*

PLEASE DUPLICATE AS NECESSARY

Rev. 2/12

Appendix 3 Teacher information letter



Researcher: Ariel Lindorff-Vijayendran, PhD Candidate, University of Oxford

Research Interview Invitation

Dear Prospective Participant:

You are invited to participate in a research study, **Support organizations and school improvement: A mixed-methods study of roles and effects of support organizations and networks in the context of New York City public schools**. As a former New York City public school teacher myself, I consider it a priority to include practitioners' voices in this research.

Purpose: To investigate the nature and impact of support organization/ network involvement in New York City public middle schools, and to explore processes and interactions between and within schools and networks.

Research Activity: An individual interview of approximately 20 minutes, to be scheduled at your convenience.

If you agree to participate or have any questions, please email me at ariel.vijayendran@education.ox.ac.uk , or call/text (646)666-3161.

Risks/Voluntary Consent: *This is a minimal-risk study. Participation is completely voluntary, and you may withdraw your consent without consequence. All information you provide will be kept confidential, and anything you say that might identify you, another individual or the school will be removed from the record. If you wish, your interview transcript will be shared with you so that you may comment, to ensure that your perspective is represented as accurately as possible. If you consent to audio recording, the digital audio file will be kept encrypted and password-protected, accessed only by the researcher. All raw data will be destroyed (securely erased) 3 years after the end of the project.*

Benefits: This research...

- Gives voice to the perspectives of multiple stakeholders in your school's network
- Contributes to the existing body of knowledge on school improvement, support and partnerships in education
- Has the potential to inform relevant policy and practice, which in turn may positively impact those working in New York City schools.

I hope that you will consider participating in this research, and contributing your insight and perspective!

Sincerely,

Ariel Lindorff-Vijayendran
Doctoral Candidate, University of Oxford, Department of Education

Appendix 4 Interview consent form



Researcher: Ariel Lindorff-Vijayendran, PhD Candidate, University of Oxford

Support organizations and school improvement: A mixed-methods study of roles and effects of support organizations and networks in the context of New York City public schools

By signing below, I confirm that:

1) I have read and understood the information sheet for the above study, have had the opportunity to ask questions about the study and received satisfactory answers.
2) My participation is strictly voluntary, and I am free to withdraw my consent at any time and for any reason I see fit, without consequence to myself.
3) I understand that this research will be written up as a thesis, which will be deposited in print and online in the Oxford University archives, likely with open access. Selected findings may also appear in future reports/articles/presentations.
4) I understand that **no identifying** information about me or any other individual will be included in any of the above, and that all confidential or identifying information (including any audio recording) will be anonymized and kept encrypted and password protected, or in a locked cabinet for hard copy items such as this form, to be accessed only by the researcher.
5) I understand that the project has been reviewed by and received ethical clearance from the University of Oxford Central University Research Ethics Committee as well as the New York Department of Education Institutional Review Board.
6) If I have a concern about any aspect of this project, I may speak to the researcher concerned (646) 666-3161) who will do her best to answer my query. If I remain unhappy and wish to make a formal complaint, I may contact:

- The NYCDOE Institutional Review Board (IRB@schools.nyc.gov)
- The Research Ethics Committee at the University of Oxford (ethics@socsci.ox.ac.uk; +44 (0)1865 614871; Social Sciences & Humanities Inter-Divisional Research Ethics Committee, Oxford University, Hayes House, 75 George Street, Oxford, OX1 2BQ, UK)

7) ***I hereby consent to be interviewed for this research.***

Participant Signature: _____

Date (month/day/year): _____-_____-_____

Do you consent to have your interview audio taped (optional)? ☐Yes ☐No
If yes, please sign again below:

Participant Signature: _____

Researcher Signature: _____ **Date:** _____

[Researcher use only] Participant ID#:

Appendix 5 Summaries of empty model results (2009-10, 2010-11, 2012-13)

The following tables (Tables 58 through 63) provide summary results from empty model comparisons for each of the years to which the analysis was extended in Chapter 7. Each table displays results of 4-classification (students within schools within networks and districts), 3-level (students within schools within networks, as well as students within schools within districts), and 2 level models. Tables include DIC values and ICC values for each level or classification.

Table 58 Summary of empty model results, Literacy 2009-10

Literacy 2009-10						
Model	**3 Level (District)**			**2 Level**		
	Mean	S.E.	95%CI	Mean	S.E.	95%CI
Fixed part						
Intercept	-0.022	0.046	(-0.111,0.068)	-0.039	0.022	(-0.081,0.004)
Random part						
Level: district	0.054	0.019	(0.027,0.099)			
Level: school	0.154	0.011	(0.133,0.178)	0.200	0.014	(0.174,0.229)
Level: student	0.827	0.004	(0.818,0.835)	0.827	0.004	(0.818,0.835)
DIC:	188206.8			188209.0		
pD:	429.9			434.0		
Units: district	32					
Units: school	453			453		
Units: student	70925			70925		
ρ (district ICC)	0.052					
ρ (school ICC)	0.149			0.195		
ρ (student ICC)	0.799			0.805		

Table 59 Summary of empty model results, Mathematics 2009-10

Mathematics 2009-10						
Model	**3 Level (District)**			**2 Level**		
	Mean	S.E.	95%CI	Mean	S.E.	95%CI
Fixed part						
Intercept	-0.059	0.056	(-0.166,0.056)	-0.080	0.024	(-0.126,-0.032)
Random part						
Level: district	0.082	0.027	(0.043,0.146)			
Level: school	0.182	0.013	(0.157,0.209)	0.251	0.017	(0.219,0.287)
Level: student	0.761	0.004	(0.753,0.769)	0.761	0.004	(0.753,0.730)
DIC:	182301.3			182303.1		
pD:	434.9			439.1		
Units: district	32.0					
Units: school	453			453		
Units: student	70925			70925		
ρ (district ICC)	0.080111					
ρ (school ICC)	0.177			0.248		
ρ (student ICC)	0.743			0.752		

Table 60 Summary of empty model results, Literacy 2010-11

Literacy 2010-11												
Model	**Cross-classified**			**3 Level (Network)**			**3 Level (District)**			**2 Level**		
	Mean	S.E.	95%CI	Mean	S.E.	95%CI	Mean	S.E.	95%CI	Mean	S.E.	95%CI
Fixed part												
Intercept	-0.030	0.046	(-0.118,0.062)	-0.048	0.031	(-0.109,0.012)	-0.030	0.045	(-0.122,0.061)	-0.046	0.021	(-0.088,-0.006)
Random part												
Level: network	0.005	0.005	(0.001,0.017)	0.027	0.012	(0.009,0.055)						
Level: district	0.050	0.017	(0.025,0.092)				0.053	0.018	(0.026,0.096)			
Level: school	0.149	0.011	(0.129,0.173)	0.176	0.013	(0.151,0.203)	0.153	0.011	(0.132,0.176)	0.199	0.014	(0.174,0.228)
Level: student	0.815	0.004	(0.806,0.824)	0.815	0.004	(0.806,0.824)	0.815	0.004	(0.806,0.824)	0.815	0.004	(0.806,0.824)
DIC:	182691.0			182693.6			182691.0			182694.3		
pD:	441.6			445.1			441.7			446.3		
Units: network	57			57								
Units: district	32						32					
Units: school	477			477			477			477		
Units: student	68023			68023			68023			68023		
ρ (network ICC)	0.005			0.027								
ρ (district ICC)	0.049						0.052					
ρ (school ICC)	0.146			0.173			0.150			0.197		
ρ (student ICC)	0.799			0.801			0.798			0.803		

Table 61 Summary of empty model results, Mathematics 2010-11

Mathematics 2010-11 Model	Cross-classified			3 Level (Network)			3 Level (District)			2 Level		
	Mean	S.E.	95%CI	Mean	S.E.	95%CI	Mean	S.E.	95%CI	Mean	S.E.	95%CI
Fixed part												
Intercept	-0.068	0.124	(-0.165,0.036)	-0.083	0.034	(-0.150,-0.017)	-0.069	0.051	(-0.173,0.034)	-0.085	0.022	(-0.129,-0.043)
Random part												
Level: network	0.004	0.003	(0.000,0.013)	0.036	0.014	(0.013,0.070)						
Level: district	0.066	0.022	(0.034,0.114)				0.069	0.023	(0.036,0.124)			
Level: school	0.163	0.012	(0.141,0.188)	0.193	0.015	(0.166,0.223)	0.165	0.012	(0.143,0.190)	0.223	0.015	(0.195,0.255)
Level: student	0.792	0.004	(0.783,0.800)	0.792	0.004	(0.783,0.800)	0.792	0.004	(0.783,0.800)	0.792	0.004	(0.783,0.800)
DIC:	180707.4			180709.5			180707.4			180709.8		
pD:	444.1			447.7			444.1			449.0		
Units: network	57			57								
Units: district	32						32					
Units: school	477			477			477			477		
Units: student	68023			68023			68023			68023		
ρ (network ICC)	0.004			0.035								
ρ (district ICC)	0.065						0.067					
ρ (school ICC)	0.159			0.189			0.160			0.220		
ρ (student ICC)	0.773			0.776			0.772			0.780		

Table 62 Summary of empty model results, Literacy 2012-13

Literacy 2012-13												
Model	**Cross-classified**			**3 Level (Network)**			**3 Level (District)**			**2 Level**		
	Mean	S.E.	95%CI	Mean	S.E.	95%CI	Mean	S.E.	95%CI	Mean	S.E.	95%CI
Fixed part												
Intercept	-0.056	0.054	(-0.165,0.048)	-0.057	0.040	(-0.134,0.024)	-0.063	0.049	(-0.162,0.031)	-0.082	0.022	(-0.126,-0.038)
Random part												
Level: network	0.017	0.010	(0.003,0.040)	0.049	0.018	(0.022,0.091)						
Level: district	0.074	0.024	(0.038,0.132)				0.077	0.025	(0.040,0.136)			
Level: school	0.150	0.012	(0.129,0.175)	0.189	0.014	(0.164,0.218)	0.162	0.012	(0.141,0.187)	0.229	0.016	(0.201,0.262)
Level: student	0.787	0.005	(0.778,0.795)	0.787	0.005	(0.778,0.795)	0.787	0.005	(0.778,0.795)	0.787	0.005	(0.778,0.795)
DIC:	173523.9			173527.3			173524.9			173529.6		
pD:	456.9			461.4			457.5			463.7		
Units: network	56			56								
Units: district	32						32					
Units: school	483			483			483			483		
Units: student	66619			66619			66619			66619		
ρ (network ICC)	0.017			0.048								
ρ (district ICC)	0.072						0.075					
ρ (school ICC)	0.146			0.184			0.158			0.226		
ρ (student ICC)	0.765			0.767			0.767			0.774		

Table 63 Summary of empty model results, Mathematics 2012-13

Mathematics 2012-13												
Model	**Cross-classified**			**3 Level (Network)**			**3 Level (District)**			**2 Level**		
	Mean	S.E.	95%CI	Mean	S.E.	95%CI	Mean	S.E.		Mean	S.E.	95%CI
Fixed part												
Intercept	-0.111	0.061	(-0.231,0.010)	-0.109	0.043	(-0.193,-0.022)	-0.110	0.056	(-0.224,-0.002)	-0.127	0.024	(-0.174,-0.079)
Random part												
Level: network	0.013	0.009	(0.001,0.036)	0.058	0.021	(0.025,0.106)						
Level: district	0.103	0.033	(0.055,0.182)				0.105	0.033	(0.056,0.185)			
Level: school	0.180	0.014	(0.155,0.208)	0.231	0.017	(0.201,0.267)	0.189	0.013	(0.165,0.217)	0.278	0.019	(0.243,0.316)
Level: student	0.736	0.004	(0.728,0.744)	0.736	0.004	(0.728,0.744)	0.736	0.004	(0.728,0.744)	0.736	0.004	(0.728,0.744)
DIC:	169069.6			169071.5			169070.0			169073.1		
pD:	462.1			466.4			462.3			468.1		
Units: network	56			56								
Units: district	32						32					
Units: school	483			483			483			483		
Units: student	66619			66619			66619			66619		
ρ (network ICC)	0.013			0.056								
ρ (district ICC)	0.100						0.102					
ρ (school ICC)	0.174			0.226			0.183			0.274		
ρ (student ICC)	0.713			0.718			0.714			0.726		

Appendix 6 Normal probability plots for model residuals: Final CVA models

The following figures (Figures 29 through 36) display normal probability plots of higher-level (school) residuals in the final models for each subject in each year. These were used as diagnostic checks, as normally-distributed residuals is a basic assumption for the multilevel analyses reported in the thesis. As the figures demonstrate, there were some indications of departure from normality, but the plots suggest that the residuals are sufficiently normally distributed for the present purposes.

Figure 22 Normal probability plot (L2), 2009-10 Literacy

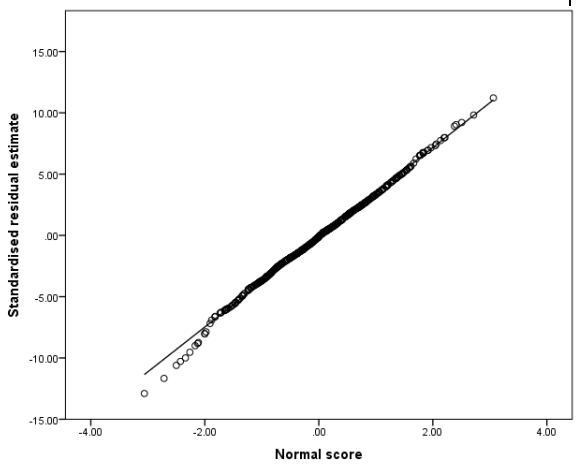


Figure 23 Normal probability plot (L2), 2009-10 Mathematics

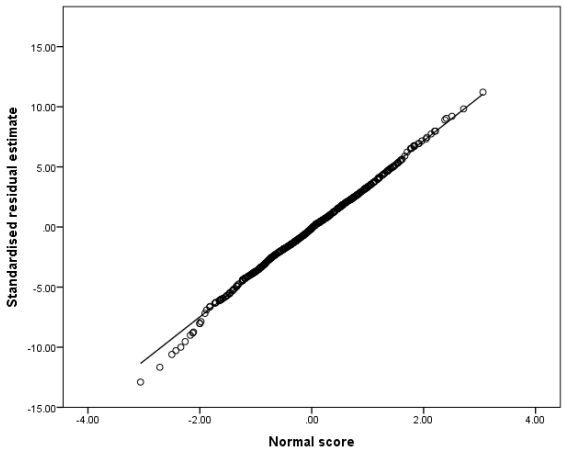


Figure 24 Normal probability plot (L2), 2010-11 Literacy

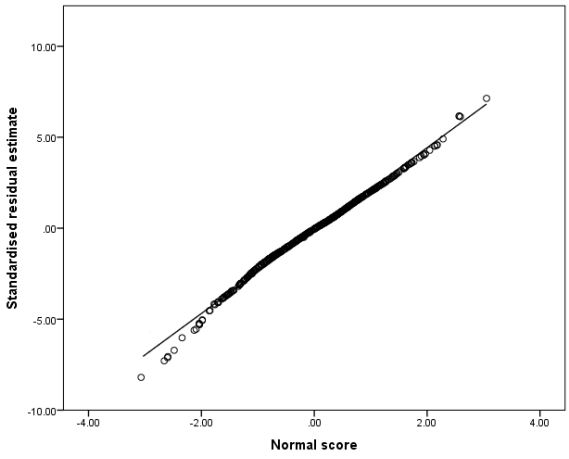


Figure 25 Normal probability plot (L2), 2010-11 Mathematics

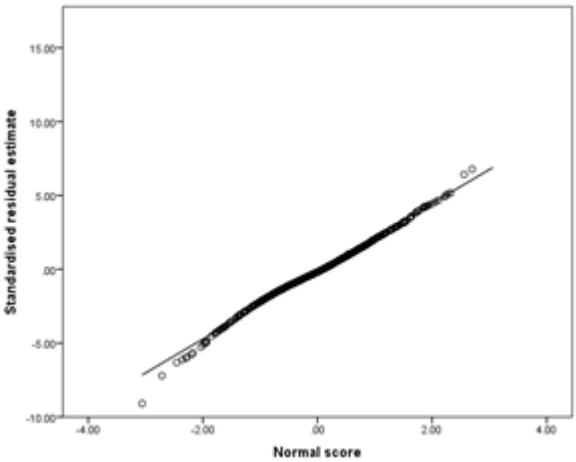


Figure 26 Normal probability plot (L2), 2011-12 Literacy

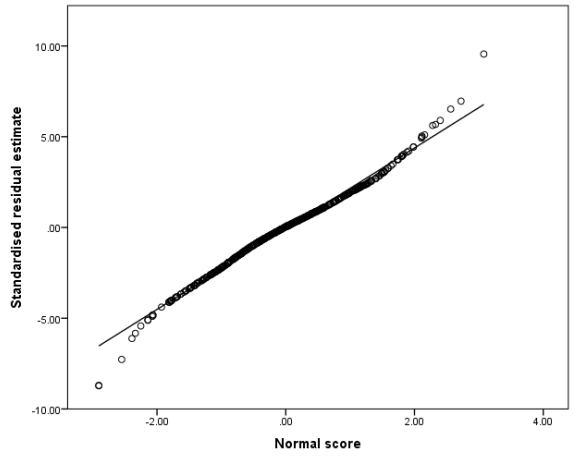


Figure 27 Normal probability plot (L2), 2011-12 Mathematics

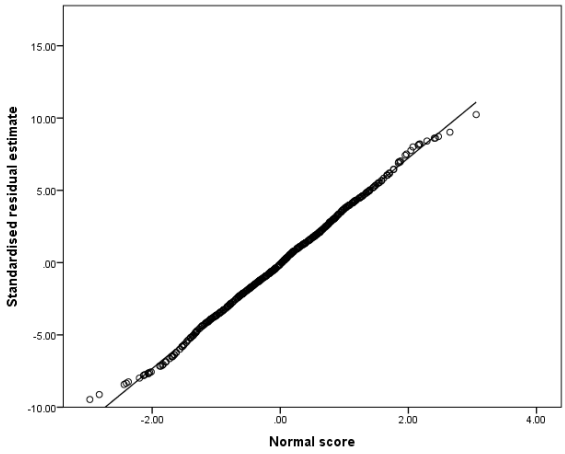


Figure 28 Normal probability plot (L2), 2012-13 Literacy

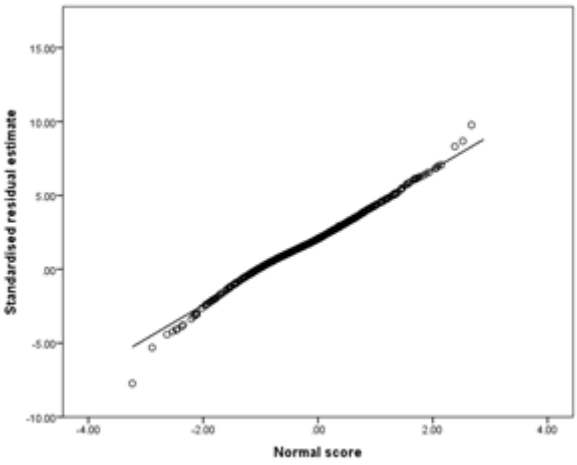


Figure 29 Normal probability plot (L2), 2012-13 Mathematics

