

Function-focused implementation fidelity for complex interventions: the case of Studio Schools

Thesis submitted for the Degree of Doctor of Philosophy in Social Intervention

Felix Constantijn van Urk, MSc
Green Templeton College
Candidate Number: 491144
Trinity Term 2016



Abstract

This thesis is concerned with an initial assessment of the implementation of Studio Schools, a novel and highly flexible model of secondary education, in England. Responding to the methodological challenges towards evaluating a ‘standard’ national social programme that is encouraged to be adapted to context by local schools, the thesis also reports the development, operationalisation, and testing of a new approach towards the concept of implementation fidelity for evaluation science.

The thesis commences by presenting the modern-historic foundations and challenges of the current English secondary education system that gave rise to Studio Schools, and describing the nature and objectives of the schools. This is followed by a discussion of the general challenges involved in the development and evaluation of complex social interventions and the specific challenges presented by the case of Studio Schools. The remainder of the thesis reports the development, use, and assessment of methods to overcome these challenges –with particular focus on evaluating implementation as part of process evaluations– as well as the current state of implementation in the schools. Delphi-inspired consensus methods were used in order to develop an explicit programme theory for Studio Schools where none previously existed, involving stakeholders in the theory specification process. The process demonstrated that stakeholders without a background in programme evaluation can agree to a specific and explicit theory of change after a programme was designed but prior to its evaluation. Next, a novel conceptual approach towards defining and measuring implementation fidelity was developed to translate a standard programme theory into flexible implementation measures. This approach focuses on the functions –or targeted change mechanisms– of a programme alongside its form of

a given set of activities. Implementation measures were developed in the form of quantitative, paper-based questionnaires that were used to rate form- and function-focused fidelity of implementation of project-based learning (PBL) and personal coaching in schools on ordinal Likert scales. These measures were piloted and refined, and subsequently tested for their psychometric properties through the use of factor analysis in addition to established methods for determining the reliability of instruments in terms of internal consistency and inter-rater agreement. Findings show that it is feasible to monitor programme functions alongside form in process evaluations, and that the validity and reliability of measures based on this approach can be established using common psychometric methods. The measures developed earlier in the thesis were used by the doctoral candidate as well as teachers and students to rate the current state of implementation practices of PBL and coaching in Studio Schools was monitored over a period of four months in four participating schools. Ratings were based on observations made in-vivo or based on video- and audio recordings made during repeated visits to the schools. Quantitative implementation scores were calculated per rater group for PBL and coaching by aggregating ratings given to individual sessions, and were compared within and between schools. Spearman's correlation coefficients were calculated to assess correlation between form- and function-focused fidelity scores. The results of this study imply that implementation in Studio Schools likely varies substantially between individual schools and can be improved in all of them, but also suggest that the model could be evaluated for its effectiveness as long as implementation and process are carefully monitored. The additions of this thesis to the evaluation literature are considered, as well as its strengths and limitations and implications for practice and research.

Acknowledgements

Completing a doctorate is not a quick and easy task. As with most accomplishments, it is also not something one does alone. On my way to writing this thesis I have been helped, supported and inspired by many people without whom I would not have been able to complete this amazing and challenging journey.

I would like to start by thanking the participants in my research project. From the experts I interviewed at the onset of the project, to the teachers and students in each of the four participating Studio Schools: I am eternally grateful for your time. My special thanks go out to the people at the Studio Schools Trust and the school principals. Without you, this doctoral project would not have existed and I would have never have learned so much from it.

Many thanks go out to Paul Montgomery and Chris Bonell for guiding me through the DPhil process and for the intellectual and emotional support they have given me during the past years. Paul, you have showed faith in me from my first months in Oxford as a Master's through my last as a DPhil, and have taught me so many valuable lessons and skills as part of my education here that I will bring with me on my next paths. Chris, you took me on as your student and helped develop the core ideas presented in this thesis from the start of the project. Despite switching institutions back to London you have stayed involved and continued to improve my thinking, reasoning, and writing throughout.

I would also like to express my gratitude to my examiners, who took the time to help sharpen my arguments and improve my analyses further.

Several fellow researchers have improved the rigour of my research by helping me improve and pilot my materials and code participant responses. Sean Grant, Elizabeth Nye, Ben Verboom, and of course Charlotte, your help has had a great impact on my work and I really enjoyed working together with you.

The quality of my work was further improved by the thoughtful and repeated feedback I received from Frances Gardner and David Humphreys. I have found your analyses of my work and suggestions for improving it of extraordinary quality and am thoroughly grateful for the time you have taken to provide them. Special thanks in this regard also go out to G.J. Melendez-Torres and Thees Spreckelsen. G.J., you are the most resourceful researcher I know, and a very dear friend. The amount of times you replaced a ‘can’t’ with a ‘can’ are countless, and you essentially taught me advanced statistics on the fly (well, advanced for me that is). Thees, your calm and clever advice has been a source of comfort in the midst of many analyses ‘crises’ (they certainly were in my head), whether conceptually or statistically.

Becoming a scientist foremost implies becoming at ease with doubt. This is done much easier when one is assured by love and friendship. In Oxford, my closest friends were there for me whenever I needed them. Sean, Jonah, and Sarah-Ann, I will forever miss our moments hanging out in the Royal Oak and the Jericho Café or rocking out to our music at the balls we played. Jenny and Alice, I must have spent as much time on your couch talking, moaning, listening, and laughing as I spent behind my computer working in the department and would never have been able to finish the thesis without you. At home in the Netherlands, the warm support of family was always enough to send me back to Oxford with the energy to get back to work. Mom, dad, and Valentijn, I know you believe in me and I believe in us. Rob, Wies, Marlou

and Stijn, I cannot tell you how blessed I feel to be able to call you family, and I look forward to adding more shared chapters.

Finally, I would not have been able to do this without you, Charlotte. You were my girlfriend, became my fiancé, and then my wife during this DPhil, and you are the most important person in my life. Wherever we go, we go together and make the best of it. I feel your love and support every day and am so happy we will spend the rest of our lives together.

Table of Contents

List of tables and figures	xi
Chapter I: Background and objectives of this thesis	1
<u>1.1 Introduction</u>	<u>1</u>
<u>1.2 Foundations and challenges of the current English secondary education system</u>	<u>2</u>
<i>1.2.1 Important post-war trends and developments within English secondary education policy</i>	2
<i>1.2.2 The current English secondary education system</i>	11
<i>1.2.3 Shortcomings of the system</i>	13
<i>1.2.4 Section summary</i>	16
<u>1.3 Studio Schools</u>	<u>17</u>
<i>1.3.1 Description and development</i>	17
<i>1.3.2 The Studio School model</i>	19
<i>1.3.3 Evaluating Studio Schools</i>	32
<i>1.3.4 Section summary</i>	37
<u>1.4 Evaluating the implementation and adaptation of interventions</u>	<u>38</u>
<i>1.4.1 Why evaluate implementation?</i>	38
<i>1.4.2 The importance of implementation fidelity</i>	40
<i>1.4.3 Adaptation of interventions</i>	44
<i>1.4.4 The tension between implementation fidelity and contextual adaptation</i>	50
<i>1.4.5 Recent developments in the fidelity-versus-adaptation debate</i>	52
<i>1.4.6 Section summary</i>	55
<u>1.5 Objectives and research strategy of this thesis</u>	<u>57</u>
<u>1.6 Chapter summary</u>	<u>60</u>
Chapter II: Specifying the programme theory of Studio Schools	62
<u>2.1 Introduction</u>	<u>62</u>
<i>2.1.1 Research question</i>	62
<u>2.2 Programme theories and stakeholder involvement</u>	<u>62</u>
<i>2.2.1 The importance of explicit programme theories</i>	62
<i>2.2.2 Elucidating implicit programme theories and building logic models</i>	65
<i>2.2.3 The Delphi method</i>	67

<u>2.3 Methods</u>	<u>69</u>
2.3.1 <i>Participants</i>	69
2.3.2 <i>Preliminary interviews</i>	69
2.3.3 <i>Delphi process</i>	70
<u>2.4 Results</u>	<u>73</u>
2.4.1 <i>Delphi process</i>	73
2.4.2 <i>Creation of the Studio Schools logic model</i>	75
<u>2.5 Discussion</u>	<u>76</u>
2.5.1 <i>Summary of results</i>	76
2.5.2 <i>Strengths of this study compared to other approaches</i>	77
2.5.3 <i>Limitations</i>	79
<u>2.6 Conclusion</u>	<u>82</u>
<u>2.7 Chapter summary</u>	<u>83</u>
 Chapter III: Approach towards monitoring implementation fidelity in Studio Schools	 84
<u>3.1 Introduction</u>	<u>84</u>
3.1.1 <i>Difficulties with defining implementation measures for Studio Schools</i>	84
<u>3.2 Revisiting implementation fidelity</u>	<u>85</u>
<u>3.3 Introducing a novel conceptualisation of implementation fidelity</u>	<u>89</u>
3.3.1 <i>What constitutes social intervention programmes?</i>	89
3.3.2 <i>Redefining programme components: activities and functions</i>	90
3.3.3 <i>The relationship between activities, functions, and outcomes</i>	93
3.3.4 <i>Distinguishing between two notions of implementation fidelity</i>	97
<u>3.4 Implications for collection and analysis of implementation data</u>	<u>98</u>
3.4.1 <i>Implications for data collection</i>	99
3.4.2 <i>Analytical implications</i>	100
<u>3.5 Discussion</u>	<u>102</u>
3.5.1 <i>Differences with previous efforts to increase the flexibility of fidelity</i>	102
3.5.2 <i>The utility of this approach for implementation and evaluation research</i>	104
<u>3.6 Chapter summary</u>	<u>107</u>

Chapter IV: Defining implementation measures for Studio Schools	109
<u>4.1 Introduction</u>	<u>109</u>
4.1.1 <i>Focus on project-based learning and personal coaching</i>	109
4.1.2 <i>Fidelity as implementation dimension of focus</i>	109
4.1.3 <i>Overview of methods and procedures</i>	110
<u>4.2 Basic theoretical framework</u>	<u>111</u>
4.2.1 <i>Procedure</i>	112
4.2.2 <i>A basic theoretical framework for project-based learning in Studio Schools</i>	112
4.2.3 <i>A basic theoretical framework for personal coaching in Studio Schools</i>	114
<u>4.3 Specifying the activities and functions of PBL and coaching</u>	<u>116</u>
4.3.1 <i>Procedure</i>	116
4.3.2 <i>Activities and functions of PBL and coaching in Studio Schools</i>	118
<u>4.4 Operationalising the activities and functions of PBL and coaching</u>	<u>122</u>
4.4.1 <i>Data sources used</i>	123
4.4.2 <i>Developing checklists and questionnaires</i>	125
4.4.3 <i>Overview of original measures</i>	128
<u>4.5 Discussion</u>	<u>134</u>
4.5.1 <i>Programme theory development process</i>	134
4.5.2 <i>Operationalisation process and types of data collected</i>	135
<u>4.6 Chapter summary</u>	<u>137</u>
 Chapter V: Assessing the validity and reliability of used measures	 139
<u>5.1 Introduction</u>	<u>139</u>
5.1.1 <i>Research question</i>	139
<u>5.2 Data collection</u>	<u>139</u>
5.2.1 <i>Settings and sampling</i>	140
5.2.2 <i>Procedure</i>	146
<u>5.3 Construct validity and internal consistency</u>	<u>147</u>
5.3.1 <i>Methods and measures</i>	147
5.3.2 <i>Procedure</i>	152
5.3.3 <i>Findings</i>	153
<u>5.4 Inter-rater agreement</u>	<u>167</u>
5.4.1 <i>Methods and measures</i>	167
5.4.2 <i>Procedure</i>	169

5.4.3 Findings	170
<u>5.5 Overview of definitive measures</u>	<u>171</u>
5.5.1 Alterations to original measures based on psychometric assessment	171
5.5.2 Definitive measures for monitoring implementation of PBL	173
5.5.3 Definitive measures for monitoring implementation of personal coaching	174
<u>5.6 Discussion</u>	<u>176</u>
5.6.1 Choice of methods for answering this research question	176
5.6.2 Operational issues faced during data collection	177
5.6.3 Modification of theory during the psychometric assessment process	179
5.6.4 PBL confirmatory factor analysis results further examined	181
5.6.5 Inter-rater agreement results further examined	184
5.6.6 Sampling issues	186
5.6.7 Further implications for function-focused fidelity monitoring	188
<u>5.7 Conclusion</u>	<u>189</u>
<u>5.8 Chapter summary</u>	<u>191</u>
 Chapter VI: Current state of implementation in Studio Schools	 192
<u>6.1 Introduction</u>	<u>192</u>
6.1.1 Research questions	192
<u>6.2. Methods</u>	<u>193</u>
6.2.1 Settings and sampling	193
6.2.2 Measures	194
6.2.3 Procedure	197
6.2.4 Data analysis	197
<u>6.3 Results</u>	<u>199</u>
6.3.1 Student sample characteristics	199
6.3.2 Implementation of project-based learning in Studio Schools	205
6.3.3 Implementation of personal coaching in Studio Schools	210
6.3.4 The correlation between form-focused and function-focused fidelity scores	213
<u>6.4 Discussion</u>	<u>216</u>
6.4.1 Current implementation fidelity in Studio Schools	216
6.4.2 The relationship between form and functions	218
6.4.3 Limitations	220
<u>6.5 Conclusions</u>	<u>224</u>

6.5.1 <i>On implementation in Studio Schools</i>	224
6.5.2 <i>On the relationship between form-focused and function-focused fidelity</i>	224
<u>6.6 Chapter summary</u>	<u>225</u>
Chapter VII: Discussion and conclusion	227
<u>7.1 Introduction</u>	<u>227</u>
<u>7.2 Summary of main results</u>	<u>227</u>
7.2.1 <i>Specification of a programme theory for Studio Schools</i>	227
7.2.2 <i>Reconceptualisation of implementation fidelity for complex interventions</i>	228
7.2.3 <i>Assessment of implementation fidelity in Studio Schools</i>	229
<u>7.3 What we already knew about the addressed topics</u>	<u>230</u>
7.3.1 <i>Programme theory development</i>	230
7.3.2 <i>The fidelity versus adaptation debate</i>	231
7.3.3 <i>Research into Studio Schools</i>	232
<u>7.4 What this thesis adds</u>	<u>232</u>
7.4.1 <i>Theory-driven evaluation, implementation fidelity, and adaptation</i>	232
7.4.2 <i>Studio Schools</i>	235
<u>7.5 Strengths and limitations</u>	<u>235</u>
<u>7.6 Implications for research</u>	<u>239</u>
7.6.1 <i>Implications for process evaluation and the monitoring of implementation</i>	239
7.6.2 <i>Implications for stakeholder involvement and participatory research</i>	242
7.6.3 <i>Implications for future research on Studio Schools</i>	243
<u>7.7 Implications for practice in Studio Schools</u>	<u>244</u>
7.7.1 <i>Ensure consistency of the Studio School approach</i>	245
7.7.2 <i>Build appropriate levels of organisational capacity</i>	246
<u>7.8 Closing remarks</u>	<u>268</u>
References	250
Appendices	285

List of tables and figures

Figure 1.1	The Studio Schools CREATE®-framework of transferable skills	20
Figure 2.1	Flow of participants and items through the Delphi study	74
Figure 2.2	The Studio Schools logic model as resulted from the study	78
Figure 3.1	A programme component comprises an activity/set of activities and its function(s)	93
Figure 3.2	An action is causal only if its outcome is triggered by a mechanism acting in context	95
Figure 3.3	Programme activities affect mechanisms of change and outcomes through their functions	96
Figure 3.4	Graphic depiction of a hypothetical relationship between programme activities, functions, mechanisms of effect, and outcomes	97
Figure 4.1	Flowchart depicting the development of implementation measures in Studio Schools	110
Figure 4.2	A basic overview of the components, mechanisms of effect, and outcomes of project-based learning as theorised for Studio Schools	114
Figure 4.3	A basic overview of the components, mechanisms of effect, and outcomes of personal coaching as theorised for Studio Schools	115
Figure 4.4	A logic model depicting the relationships between activities, functions, mechanisms of effect, and outcomes of project-based learning as theorised for Studio Schools	120
Figure 4.5	A logic model depicting the relationships between activities, functions, mechanisms of effect, and outcomes of personal coaching as theorised for Studio Schools	122
Figure 5.1	Flowchart depicting analysis procedure for determining the validity and internal consistency of used measures	153
Figure 5.2	Scree plot suggesting a factor structure with three factors for the student PBL questionnaire	156
Figure 5.3a	Overview of factor structure used to analyse PBL implementation data, estimated using WLS estimation procedures, using initial indicator names	162
Figure 5.3b	Overview of factor structure used to analyse PBL implementation data, estimated using WLS estimation procedures, using revised indicator names	163
Figure 6.1	Flowchart depicting the development of implementation measures for Studio Schools	195
Figure 6.2	Flowchart depicting analysis procedures for determining the validity and internal consistency of used measures (Copied from Fig. 5.1)	195
Table 1.1	Overview of State Schools in England	12
Table 2.1	A7S Criteria for the classification of items after Delphi Round Two ratings by participants	72
Table 2.2	Classification of items based on ratings in Delphi Round Two	75
Table 2.3	Example of final decision process on inclusion and exclusion of items	76
Table 3.1	Distinction between simple, complicated, and complex interventions	85

Table 3.2	Examples of standardisation by form and function as defined by Hawe et al., 2004a	103
Table 3.3	A systematic approach towards modifying programme theories based on anticipated or demonstrated contextual influences	107
Table 4.1	Overview of the activities and functions of each component of project-based learning as theorised for Studio Schools	118
Table 4.2	Overview of the activities and functions of each component of personal coaching as theorised for Studio Schools	121
Table 4.3	Overview of original scales and items used for measuring implementation of project-based learning lessons in Studio Schools	128
Table 4.4	Overview of original scales and items used for measuring implementation of project-based learning structuring in Studio Schools	130
Table 4.5	Overview of original scales and items used for measuring implementation of personal coaching sessions in Studio Schools	130
Table 4.6	Questionnaire composition by type of item per category of rater: project-based learning	133
Table 4.7	Questionnaire composition by type of item per category of rater: personal coaching	133
Table 4.8	Operationalisation example: PBL function FC1	133
Table 5.1	Overview of sessions, students, and teachers included per school in PBL data collection	144
Table 5.2	Overview of gender and year groups of students included per school in PBL data collection	144
Table 5.3	Overview of gender of teachers included per school in PBL data collection	144
Table 5.4	Overview of sessions, students, and teachers included per school in coaching data collection	145
Table 5.5	Overview of gender and year groups of students included per school in coaching data collection	145
Table 5.6	Overview of gender of coaches included per school in coaching data collection	146
Table 5.7	Overview of analysis procedures used to determine the construct validity of used measures	151
Table 5.8	Overview of analysis procedures used to determine the internal consistency of used measures	152
Table 5.9	Internal consistency statistics for original scales and subscales used to monitor project-based learning in Studio Schools	154
Table 5.10	Recoding/removal of PBL items poorly correlated with overall structure in student questionnaire	155
Table 5.11	Factor loadings after orthogonal varimax rotation for three factors for the student PBL questionnaire	156
Table 5.12	Factor loadings after orthogonal varimax rotation for four factors for the student PBL questionnaire	157
Table 5.13	Fit statistics for models that provided significant fit with student-reported PBL data, using WLS estimation	159
Table 5.14:	Parameter estimates for Model 1	159
Table 5.15	Parameter estimates for Model 2	160

Table 5.16:	Internal consistency statistics for revised scales and subscales used to monitor project-based learning in Studio Schools upon completion of factor analyses	164
Table 5.17	Internal consistency statistics for original scales and subscales used to monitor personal coaching in Studio Schools	165
Table 5.18	Recoding/removal of personal coaching items poorly correlated with overall structure in student questionnaire	166
Table 5.19	Internal consistency statistics for revised scales and subscales used to monitor personal coaching in Studio Schools	167
Table 5.20	Overview of analysis procedures used to determine the inter-rater reliability of used measures	169
Table 5.21	Inter-rater agreement statistics for original and revised measures used to monitor project-based learning in Studio Schools	170
Table 5.22:	Inter-rater agreement statistics for original and revised measures used to monitor personal coaching in Studio Schools	171
Table 5.23	Overview of definitive scales and items used for measuring implementation of project-based learning in Studio Schools	173
Table 5.24	Overview of definitive scales and items used for measuring implementation of personal coaching in Studio Schools	174
Table 6.1	Overview of sessions, students, and teachers included per school in PBL data collection (Copied from Table 5.1)	193
Table 6.2	Overview of sessions, students, and teachers included per school in coaching data collection (Copied from Table 5.4)	194
Table 6.3	Numeric properties of PBL scales, shown per rater group	196
Table 6.4	Numeric properties of personal coaching scales, shown per rater group	197
Table 6.5	Number of students who completed the baseline questionnaire	200
Table 6.6	Overview of students' demographic backgrounds	201
Table 6.7	Most important reasons for choosing a Studio School, as reported by students who completed the baseline questionnaire	205
Table 6.8	Project-based learning fidelity scores as defined by form, shown per rater group	206
Table 6.9	Project-based learning fidelity scores as defined by function, shown per rater group	207
Table 6.10	Personal coaching fidelity scores as defined by form, shown per rater group	212
Table 6.11	Personal coaching fidelity scores as defined by function, shown per rater group	212
Table 6.12	Correlation between aggregated form-focused and function-focused fidelity scores for project-based learning, shown per rater group	214
Table 6.13	Correlation between aggregated form-focused and function-focused fidelity scores for personal coaching, shown per rater group	214

Chapter I: Background and objectives of this thesis

1.1 Introduction

This chapter presents the context of this doctoral thesis in terms of historical and current societal trends in English secondary education as well as the scientific literature on implementation and intervention science to which it aims to make a contribution. The first section of this chapter offers a description of how the English secondary education system of today has developed since the end of the Second World War, and the difficulties it faces with regards to the education of young people. In the second section, the social programme of focus within the thesis is introduced in the form of Studio Schools. Studio Schools are designed to tackle the twin challenges of a shortage of transferable employability skills of secondary school leavers and low levels of engagement with school experienced by a large group of young people. The development of the model as well as its components is discussed, and a particular focus is given to the challenges towards evaluating the model. The third section of the chapter is concerned with the field of implementation science. The section concentrates around two important concepts within this field: implementation fidelity and the adaptation of interventions. Traditionally, these concepts have been presented as in tension with one another, as the need for implementers to adapt innovations to their specific circumstances is held to conflict with the need to implement such innovations as they were intended and tested for their effectiveness. Finally, the objectives of this thesis and the research questions it seeks to answer are presented in the fourth section.

1.2 Foundations and challenges of the current English secondary education system

The first section of this chapter describes the English secondary education system, which provides the context for the foundation of Studio Schools. It starts with addressing some important themes that have emerged since the 1950s that have had a profound influence on the current state of the English secondary education system. As the first Studio Schools opened in 2010¹, themes are mostly described prior to the formation of the Coalition government. The section continues with a succinct overview of the current state of the system, upon which it completes with a discussion of some of the difficulties it faces.

1.2.1 Important post-war trends and developments within English secondary education policy

I. Diversity versus comprehensiveness (1944 – 1997)

The Education Act of 1944: the foundation of modern English secondary education

The foundations for the current English secondary education system were laid at the end of the Second World War in the Education Act of 1944, which practically replaced almost all previous education legislation in England and Wales (Gillard, 2011). The fundamental idea underlying the Act was that education is a continuous process consisting of stages. Up to the lawful leaving age (which was raised to fifteen), students would pass through two stages of education: primary education (until age eleven), and secondary education (age eleven to fifteen) (McCulloch, 1994). An important part of the act was to establish three main types of state-funded secondary schools to be defined as ‘Grammar Schools’, ‘Technical Schools’, and

¹ Studio Schools will be discussed in detail in Section 2 of this chapter.

‘Modern Schools’ (the last would come to lay at the heart of the development of what is now known as Comprehensive Schools). This system of three types of schools has since been known as the ‘*Tripartite System*’, although some recognise it as de facto a bipartite system due to the fact that hardly any Technical Schools were ever erected (see part III of this section on the curricular focus of state-funded secondary schools). The Act ruled that selection of students into different schools was to be based on an examination when students were aged eleven-plus in which attainment in English and Mathematics and general intelligence were assessed. Overall, the system was based on the influential assumption that intelligence was the most important factor of the mind (Chitty, 2009).

1960-1980: momentum and critique of progressive Comprehensive Schools

In the 1960s, the selection of students into schools of varying quality on the basis of the eleven-plus exam became subject to major criticism. Fed by evidence suggesting that examinations were seriously flawed and the strong inequality in facilities and prestige between Grammar and Secondary Modern schools, the movement to abolish this selection process and therefore the distinction these types of schools was gaining momentum (Chitty, 2009; Sandbrook, 2006). This movement resulted in Harold Wilson’s first Labour administration to distribute ‘Circular 10/65’, in which it declared an intention to end the selection of students into separate type of schools at eleven-plus and in which it requested local authorities to reorganise secondary education along comprehensive lines. Although welcomed on the left as a step towards greater equality, some local authorities as well as many parents were much against the prospect of losing their high-quality grammar schools in return for a more equal system that had yet to prove itself (Sandbrook, 2006). In a decade spanning between 1965 and 1975, virtually every secondary school in Wales and Scotland was

turned into a comprehensive school, with figures for England approximating 90 percent (Pring & Walford, 1996). Conflicts in opinions increased during the 1970s, when influential groups in society expressed their concerns with the output of comprehensive schools. This was testified by the well-received (particularly in the right-wing press) series of five Black Papers published by a group of influential activists between 1969 and 1977 (Cox & Boyson, 1975, 1977; Cox & Dyson, 1969a, 1969b, 1970). These criticised the ‘progressive’ teaching methods of Modern Secondary Schools which in their eyes focused too little on teaching core subject knowledge, and attacked the concept of all ability schools that prevented able students from reaching their potential. In addition to this, the coercion of Grammar Schools to quit their eleven-plus selection forced many of these to become independent (private) schools, which actually contributed to an increase in inequality within the education system (Sandbrook, 2012).

1980-1997: selection by specialisation

Not much was done under the administrations of Margaret Thatcher with regards to the process of selecting students into different state-funded schools. As part II (on control and accountability) of this section demonstrates, their policies were much more focused on the centralisation of control over the state-school curricula and the erosion of Local Education Authority (LEA) control. When John Major succeeded Thatcher in 1990, one of his key education policies was that of selection by specialisation (Chitty, 2009). This implied that schools were as much as possible to specialise in particular fields and that selection would lie in the hand of parents (not schools) who would select the schools that ‘cater for their [child’s] particular interests and aptitudes’ (Department for Education, 1992) This was rooted in the Major governments’ criticism of the comprehensive education system that it presupposed

‘that children are all basically the same and that all local communities have essentially the same educational needs’ (Department for Education, 1992). The result was that when the Conservatives left office in 1997, England had 164 Grammar Schools that together with 1155 ‘opted-out’ schools (centrally funded schools independent of local authorities) accounted for 19.6 percent of pupils in secondary schools. In addition, there were fifteen City Technology Colleges², 30 colleges specialising in language and 151 new colleges specialising in technology (but not adhering to the format and statutory basis of CTCs) (Chitty, 2009). The system was thus indeed diverse and divided.

II. Central versus local control (1980 – Present)

1980 – 1997: increased centralised control during the Conservative era

It was with the Education Reform Act of 1988 that the Thatcher-led Conservatives left their biggest mark on English education. The main aims of the act for the secondary education system included the establishment of a National Curriculum and the strengthening of central governmental power over schools (Gillard, 2011). The introduction of the National Curriculum and the General Certificate of Secondary Education (GCSE) put the content of curriculum taught at state schools directly under control of the central government (Leonard, 1988). The control of Local Education Authorities was further eroded by the opportunity the act provided for schools to opt-out of LEA control and be maintained and subject to direct control by the government (Lawton, 1989). John Major’s administration further strengthened its control over schools by bringing the inspectorate back under central governmental control with the Education (Schools) Act of 1992. This inspectorate has since then become known as the Office for Standards in Education (OFSTED).

² See part III of this section on the curricular focus of state-funded secondary schools

1997-2010: central accountability and state-civil shared control under New Labour

When Tony Blair's New Labour came into power, it aimed to keep the middle class attached to its policies and the state system. For this, it was essential to uphold the principles of parental choice and competition in secondary education that had been built up under the previous government (Glennerster, 2000). The Conservative policies of central control over schools through the National Curriculum and Ofsted and increased accountability of schools through the publication of national tables of examination performance were also retained. To further increase school and teacher accountability, the Education Act of 2002 set floor targets for students' performance in secondary schools, irrespective of their different catchment areas (Education Act, 2002). One of New Labour's main concerns was the underperformance of a large fraction of comprehensive schools in specific local (mostly urban) areas. In a response, the Learning and Skills Act of 2000 included the legal creation of the 'City Academy' –generally referred to as 'Academies' (Learning and Skills Act, 2000). These were designed as public/private partnerships in which businesses, churches, or voluntary groups would build and manage schools that were outside of the control of LEAs but would be funded and held accountable by the central government (Chitty, 2009). This policy was further built on with the 2006 Education and Inspections Act. While most Academies (nine out of the assessed eleven) came in the bottom 200 school in England when compared on test results for fourteen year olds in English, maths and science, the 2006 Education and Inspections Act still recommended that all comprehensive schools should become trust schools (not necessarily academies yet in the legal sense, but based on the same principle of community ownership of schools). Such schools were to form links with external partners (such as businesses) who

would be able to appoint the majority of the schools' governors, own their own assets, employ their own staff, and set their admission arrangements (Gillard, 2011).

2010: further emphasis on local civil control under the Coalition Government

In 2010, the Conservative-Liberal Democrat coalition led by David Cameron announced that it would further increase opportunities for civil parties such as parents, teachers, charities and local communities the chance to set up new schools. An important step towards this aim has been taken with the Academies Act of 2010 that introduced the concept of the Free School. Free Schools are 'all-ability state-funded schools set up in response to what local people say they want and need in order to improve education for children in their community' (Department for Education, 2013a). They are in effect start-up academies (Hatcher, 2011). The aim of these schools is to help raise standards of attainment, particularly for pupils from socially disadvantaged backgrounds, by offering local communities the opportunity to found state-funded schools of high quality. As Cameron was quoted in *The Guardian* in January 2011:

"For too long in our country, exercising choice to escape poor schools has been available to the richest, who could just opt out and go private, or to the middle classes who could move house to a better area, but the poorest have had to take what they're given. Not anymore." (David Cameron, quoted in Doward, 2011).

The concept of the Free School is rooted primarily in three previous school initiatives that, according to the Conservative Party have been successful: the Labour Party's academies and two marketised supply-side reforms in school systems: Free Schools in Sweden and Charter Schools in the US (The Conservative Party, 2010). However, a review of the literature on these initiatives shows that the evidence regarding the success of these models, particularly for disadvantaged young people, is mixed

(Hatcher, 2011). In addition to introducing Free Schools, the Academies Act of 2010 allowed all remaining state schools to become Academies.

III: The curricular focus of state-funded secondary schools (1944 – 2010)

The Education Act of 1944: Tripartite or Bipartite?

The Education Act of 1944 aimed to lay the foundations for a broad curriculum to be offered by the tripartite system. Whereas Grammar Schools would provide the most gifted students and Secondary Modern Schools would deliver general academic education, Technical Schools would prepare students for the vocational world. However, local education authorities generally were reluctant to spend the money needed to develop Technical Schools, which may have been due to confusion about the exact function of these schools or to the expensive equipment these required. This would set the tone for an underrepresentation of vocational education in the curricula offered by state schools in England throughout the 20th century, despite returning attempts to change this situation.

1960 – 1980: concerns for employability

As discussed earlier, Comprehensive Schools became subject to wide criticism from educational activists, local authorities and parents after their initial warm welcome in the early and mid sixties. Criticism also came from industry. A large survey of employers' attitudes to young workers conducted by the National Youth Employment Council in 1974 found that the majority of participants were becoming increasingly concerned with the employability of secondary school leavers (NYEC, 1974). Some leading industrialists argued that the shortage of skilled and adaptable employees could be attributable to problems with the education system. Most notably, the lack of industrial experience of teachers and the insufficient engagement of employers in the

planning of vocational training programs were cited as the cause of young people's poor employability (Chitty, 2009). Some of the educational reforms of the 1960s – particularly the introduction of progressive teaching methods and of comprehensive schools- were held responsible for a decline in general standards and basic skills in the mid-seventies, and for the growing gap between what young people learned in school and what they needed in the workplace (CCCS, 1981). It needs to be noted here that no compelling evidence ever existed for the strong belief that it was the deficiencies of the young (rather than a problem of the quality of work on offer) that accounted for the growing numbers of workless youth (Mizen, 2004). Furthermore, the belief that a direct relationship exists between the amount of education in a society and its economic growth rate, and that governments can fine-tune education expenditures to maximise that self-same growth rate can be questioned altogether (Wolf, 2002). Nevertheless, these beliefs were widespread in the late 1970s and caused a large movement demanding a greater control of schooling by 'non-teachers', and a reinforcement of selection at the secondary level (Chitty, 2009).

1980 – 1997: vocational curricular initiatives during the Conservative era

The first two Thatcher administrations remained cautious with regards to *implementing* any radical or innovative education policies, with the start of the Technical and Vocational Education Initiative (TVEI) by Sir Keith Joseph as a notable exception (Chitty, 2009). This initiative aimed to bring a significant element of technical and vocational education into the school curriculum and to explore and test new ways of organising and managing the education of 14-18 year-old young people (Williams & Yeomans, 1994). It was introduced through fourteen pilot projects and involved approximately 600 schools and colleges, but was never really able to secure the coherent, employability skills-related curriculum that had been

intended (and was eventually abolished in 1997) (Richardson et al., 2010). Apart from establishing the (cognitively oriented) National Curriculum, the Education Reform Act of 1988 also created the City Technology Colleges. Their purpose was to offer an option of schooling that had more direct relevance to post-school employment. However, in line with previous crossovers between academic and vocational education, this school option did not become very popular. In the end, only fifteen got off the ground, and most were later converted into 'Academies' (see section 1.1.2) (Chitty, 2009).

1997-2010: returning concerns for employability under New Labour

New Labour's strong focus on improving the academic standards of underperforming secondary schools and the methods for implementing this focus (strict monitoring of performance and the stimulation of foundation of Academies) faced some major criticism. However, the government seemed reluctant to respond to such critics. Most notably, a working group led by former Chief Ofsted Inspector Mike Tomlinson identified the UK's poor record of keeping teenagers in school and the poor status of vocational courses and qualifications as some of the main problems in secondary education (Working Group on 14-19 Reform, 2004). The recommendations for reform of the secondary education curriculum included -amongst others- the introduction of wider curricular activities (such as paid and voluntary work) and the reorganisation of coursework into a single extended project. Although the Chief Inspector and the Head of Qualifications, and heads of schools, backed these recommendations, the government rejected most of them. In addition and perhaps unintentionally, the Academy movement also meant the effective end of the City Technology Colleges. Although Academies were primarily intended to address the problem of low achieving locally controlled urban comprehensive schools, most

CTCs were also converted to Academies in order to reduce costs to school sponsors, which resulted in a deviation from their original aims.

Summary

During the second half of the twentieth century and the first decade of the twenty-first, most of the foundations were laid for the current provision of secondary education in England. These foundations are mostly rooted in the beliefs that all students should have an education of sufficient standard to progress towards their desired career and life paths, and that provision of school- and education types should be diverse enough to facilitate the needs of diverse communities and students. However, these beliefs have not been (and are not currently) shared universally, and ideas regarding how to control a system that embraces diversity but that also offers optimal opportunities for all students have differed within and between time eras. In tackling these challenges, policy makers have traditionally focused on the cognitive and academic aspects of education in the form of selection by cognitive ability, curriculum control, and setting performance standards for state schools. In these efforts, appeals throughout the past 70 years expressed by many influential civic parties for secondary education to also address vocational-, employability- and other non-cognitive skills appear to have been mostly unsuccessful.

1.2.2 The current English secondary education system

Current secondary education provision in England

Rooted in the belief that diversity in provision is key to success, the current system of secondary education in England distinguishes between a number of different types of schools. These can be divided primarily between *state schools* and *independent schools*. The core difference between these two types of schools is that the former are

funded (partly) by the UK government and are free to attend, whereas the latter are funded privately and charge fees for attendance. Independent schools are also mostly autonomous from governmental control (less regular inspection) while state schools are subject to more significant control. The vast majority of students aged between eleven and nineteen attend state schools with numbers of approximately 3.2 million versus 320 thousand students in independent schools (Department for Education, 2015). These state schools can be further divided into different types of schools. This can be done across three dimensions: 1) how and by whom schools are (partly) funded and controlled, 2) the focus of the education offered by the schools and 3) the schools' admissions policies. In the context of this doctoral thesis, only dimensions two and three are of relevance. Table 1.1 gives an overview of the types of state schools that currently exist in England in terms of their curricular focus.

State Schools in England					
Academies	City Technology Colleges	Faith Schools	Free Schools	Grammar Schools	Maintained Comprehensive Schools
- Partly state-funded all-ability schools, out of LA control. Do not have to follow the national curriculum and can set their own term times. - Require following state rules on admissions, special education needs. - Run by an academy trust that employs staff. Partly sponsored privately, with sponsors being responsible for improving and maintaining school performance. - Include University Technical Colleges, which are governed by Universities and specialise in engineering, construction, and IT.	- Partly state funded all-ability schools. Required to teach the national curriculum. - Specialise in technological and practical skills. - Partly owned by private sponsors	Can take the form of any other state school, but associated with a particular religion or faith	- Fully state-funded all-ability schools, but out of LA control. Not required to teach the national curriculum. - Can set pay and conditions for staff independently and change the length of school terms and days. - Can be set up by any social group (e.g. teachers, parents) in response to local needs	- Fully state-funded schools within LA control that hold a selective admissions policy based on cognitive ability at age 11. Require teaching the national curriculum. - Some grammar schools have converted to academies but operate de facto as grammar schools (retain their admissions policy).	Fully state-funded all-ability schools either controlled by central or local authority. Require teaching the national curriculum (number includes Community Schools, Voluntary-aided Schools, and Foundation Schools).
2075 [1]	3 [1]	639 [1]	122 [1]	164 [2]	959 [1]

Table 1.1: Overview of State Schools in England

[1] Number of schools open per January 2015: Department for Education, 2015

[2] Number of schools open per December 2012: Bolton, 2013

Table 1.1 reveals England's diverse and complicated system of state-funded secondary education. Additionally, it becomes clear that the dominance of cognitively

focused secondary schools that emerged since the 1950s is still present (even though a number of academies and free schools also focus on non-cognitive and vocational skills).

1.2.3 Shortcomings of the system

In spite of the great diversity in opportunities for secondary education offered by the English system, a considerable body of evidence shows this may not appropriately serve the needs of a large group of young people in secondary schools. Whereas some difficulties manifest themselves while students are (or should be) in school, others arise after they have left school.

Poor engagement and attainment while students are in school

The large proportion of youth in England who are disengaged with their education forms a major concern to policymakers, teachers, and other stakeholders. Although variations in definition occur, it is estimated that between one fifth and one third of all young people aged 14-16 in the United Kingdom fail to reach their academic potential as a result of lost interest in learning (Steedman & Stoney, 2004). Defined as young people who had much lower aspirations, disliked school and were far more likely to play truant, disengaged students recently represented 12% of young people in Year 9, 19% in Year 10, and 20% in Year 11 in the UK (Ross, 2009). A significant proportion of students fail to finish secondary education, as one child in five in England leaves school without a qualification with a Level 2 equivalent (Marshall, 2013). Longitudinal research by the Centre for Research on the Wider Benefits of Learning (Institute of Education, London) indicates that the decline in wellbeing in and engagement with school is strongest during early adolescence, and that there are important inequalities between students. The Centre reports that boys demonstrate

less positive feelings in school than girls, and have a stronger decline during adolescence. Furthermore, young people from low income families have less school wellbeing and engagement than others, and boys are subject to additional risk factors including coming from a single-parent household, maternal depression and alcoholism, living in social housing, and having special educational needs (Morrison Gutman et al., 2010). These findings suggest that disengagement with education is a particular problem for young people –especially boys– who already face a range of other social difficulties. Recent national attainment statistics suggest that ethnicity may place specific groups of youth at further risk, with students from a White British, African, or Caribbean background underperforming compared to their peers with other backgrounds (DfE, 2014). A strong lack of engagement with education can have serious consequences. Difficulties associated with disengagement from education reach beyond failure to reach one's potential in the long term. Educational disengagement has been associated with adverse health outcomes such as higher rates of teenage pregnancy (Bonell et al., 2003; 2005; Manlove, 1998) and higher rates of substance (mis)use (Beinart et al., 2002; Bond et al., 2007), and with poor labour market opportunities (McIntosh & Houghton, 2005). Furthermore, leaving secondary school before completion (dropping out), an important potential result of disengagement, has also been related to increased crime rates (Feinstein, 2002; Lochner & Moretti, 2001).

Perhaps related to high rates of disengagement within secondary education is the United Kingdom's suboptimal ranking when compared to international standards of attainment in secondary education. The most recently reported mean PISA scores for mathematics, reading, and science place the UK below most high-income countries in

Europe³ and place it 26th amongst all OECD countries (OECD, 2013a). Eighteen percent of fifteen year olds in the UK failed to score level 2 or higher on these tests (Marshall, 2013). Relative deprivation⁴ does not account for this number, as the UK ranks ninth highest within the OECD on this measure (Amadeo & Marshall, 2013). The excessive focus on monitoring and improving academic attainment that reached its peak under New Labour may be one of the main reasons why poor attainment in secondary schools has remained relatively unexposed. A comparative analysis of England's education system that was recently compiled by the OECD contrasts the continuous rise in attainment as measured with domestic metrics (such as the GCSE) with the lack of progress in international measures (Braconier, 2012). This suggests that an over reliance on test results in a 'high-stakes' system may encourage grade inflation, particularly as many stakeholders involved in marking students' work are strongly dependant on the results they produce (Amadeo & Marshall, 2013).

Difficulties after students have left school

Some students who stay in secondary school and receive their qualifications may still fail to gain the skills needed to succeed in the labour market, which further adds to their difficult position within it. Between 2005 and 2010, unemployment among young people (aged 15-24) in the UK increased by 8.6% versus 4.2% amongst people aged 25-54, whereas the OECD average was 3.2% and rates in countries such as Germany, the Netherlands, and Finland decreased (OECD, 2013b). Despite the rise in participation in education and training by 16 to 18 year olds since 2003, the number of youth not in education, employment or training (NEET) has been steady over this period (Department for Education, 2010). It is clear that factors beyond young

³ Such as Switzerland, the Netherlands, Finland, Belgium, Germany, France, and Denmark

⁴ The deprivation of wealth and social status that is considered the norm in the society to which an individual or group belongs and in comparable societies

people's skills affect this situation (Mizen, 2004; Wolf, 2002). Nevertheless, there is reason for concern regarding their employability. Some argue that a focus on a wider set of skills than those that are taught in secondary schools may offer part of the solution to this problem. Pioneering work by Bowles & Gintis (1976; 2002) and Heckman & Rubinstein (2001) argued that Western education systems focus strongly on cognitive skills (such as IQ) with too little attention paid to what can be called non-cognitive skills (skills such as perseverance and dependability). When controlling for cognitive ability, Heckman and Rubinstein showed that people with poor non-cognitive skills function less well in the workforce. Despite this evidence, Heckman (2006) argues that major public educational policy initiatives still measure their success based primarily on cognitive test scores. Other authors also stress the importance of stimulating non-cognitive skills proactively in schools (for example: Lleras, 2008; Rothstein, 2004). As a considerable portion of young people do not progress into higher education, it can be argued that secondary schooling has an important role to play in preparing them for the workforce.

1.2.4 Section summary

The great variety of education policy measures that have accumulated over the past 70 years has provided England with a very diverse and equally complex secondary education system. Although the wide range of ways and parties by which different school types are controlled is in line with the demand for diversity in society and resolves many practical issues, some important shortcomings of the system remain. In particular, governments' strong focus on monitoring and improving schools' performance using standardised cognitive measures throughout the past decades has resulted in a deficit of attention to the development of other important skills. These so called 'non-cognitive' or 'soft' skills (e.g. communicating, empathising, and

managing expectations) are recognised by some as equally important for students' readiness for the workforce and adult life upon leaving schools, but are typically not nurtured in English schools. In addition, in spite of their diverse management arrangements, the diversity the curricula offered by the schools appears insufficient to facilitate the needs of a large proportion of young people in the country. This is testified by numbers confirming the existence of a large group of young people who strongly lack engagement with their school and schooling, and the poor attainment in UK secondary schools when compared to other high-income countries.

1.3 Studio Schools

1.3.1 Description and development

Studio Schools –named after the Renaissance concept of the ‘studio’ in which working and learning co-occurred simultaneously– form a recently developed innovation in secondary education aimed at tackling some of the challenges found in the UK education system. The schools started operating as free schools (state funded schools set up in response to the specific education needs of children in local communities) but many have converted to or have been founded as academies. Studio Schools provide secondary education at the level of Key Stage 4 and beyond for young people of all ability between the ages of fourteen and nineteen years old, and offer a contextualised approach to learning centred around enterprise projects and students gaining work experience (Studio Schools Trust, 2012a).

Targeted problems and population

Developed as a response to some of the shortcomings of the British education system, Studio Schools aim to tackle two challenges (Studio Schools Trust, 2012b):

1. Youth disengagement with secondary education: a large number of young people are poorly engaged with school and their education. This may not only limit their ability to obtain their appropriate qualification and fulfil their potential, but may also cause them to leave school prematurely which is an important predictor of a range of adverse social and health outcomes.
2. A shortage of transferable employability skills amongst completers of secondary education. Surveys from the United Kingdom and the United States showed that the majority of employers in the samples felt that youth were insufficiently prepared for employment (CBI, 2010; ETS, 2007), which they thought was caused largely by their insufficient transferable non-cognitive skills.

Studio Schools aim to offer a new mainstream option of education (Studio Schools Trust, 2012a) They have no unique defined set of admission criteria beyond the general criterion required by free schools that these be founded and supported by a 'suitable sponsor' (to be approved by the Secretary of State for Education) and are open to students of all abilities. Students self-select into the schools for a variety of potential reasons. However, from the primary objectives listed above it is clear that the schools target youth who have become disengaged with conventional secondary education or whose needs are better fulfilled in a non-traditional, contextualised pedagogic setting.

Development and current status of the Studio Schools movement

Originating from a partnership between two organizations concerned with young people's wellbeing and education (the Young Foundation and Edge), the first Studio Schools were opened in September 2010. Currently, there are 40 Studio Schools in operation in England, with numbers increasing steadily. The schools are endorsed and fully funded by the UK Department for Education, and receive ongoing support by an overarching organisation that connects them with each other as well as partner organisations across the country (Department for Education, 2013b; Studio Schools Trust, 2013a). Schools are set up by local communities with the backing of local

businesses and employers. These resources and features place Studio Schools in the position of having the potential of achieving significant social impact.

1.3.2 The Studio School model

Like conventional English secondary schools, Studio Schools teach the national curriculum and offer key academic and vocational qualifications. All schools are found and led locally, and the qualifications offered by individual schools depend on local circumstances. However, all schools deliver qualifications at a minimum of Level 2 (5 GCSEs in English language and literature, Mathematics, and Double Award Science) (Studio Schools Trust, 2013b). In addition, all activities taking place in Studio Schools should be underpinned by a set of transferable employability skills referred to as the CREATE®-skills (Studio Schools Trust, 2013c), which is displayed in Figure 1.1. Studio Schools' students are expected to have mastered these skills to a sufficient level upon finishing their education in the schools, although they are currently not assessed across schools in a standardised manner.

Programme theory

Studio Schools currently do not work with an explicit and coherent theory of change: a theory that describes a causal chain of activities intended to produce a positive outcome (Fraser et al., 2009). As will be discussed in the final section of this chapter, the development of such a theory forms the first part of the research project reported in this thesis. However, the Studio Schools Trust does currently offer a description of its model and its components.

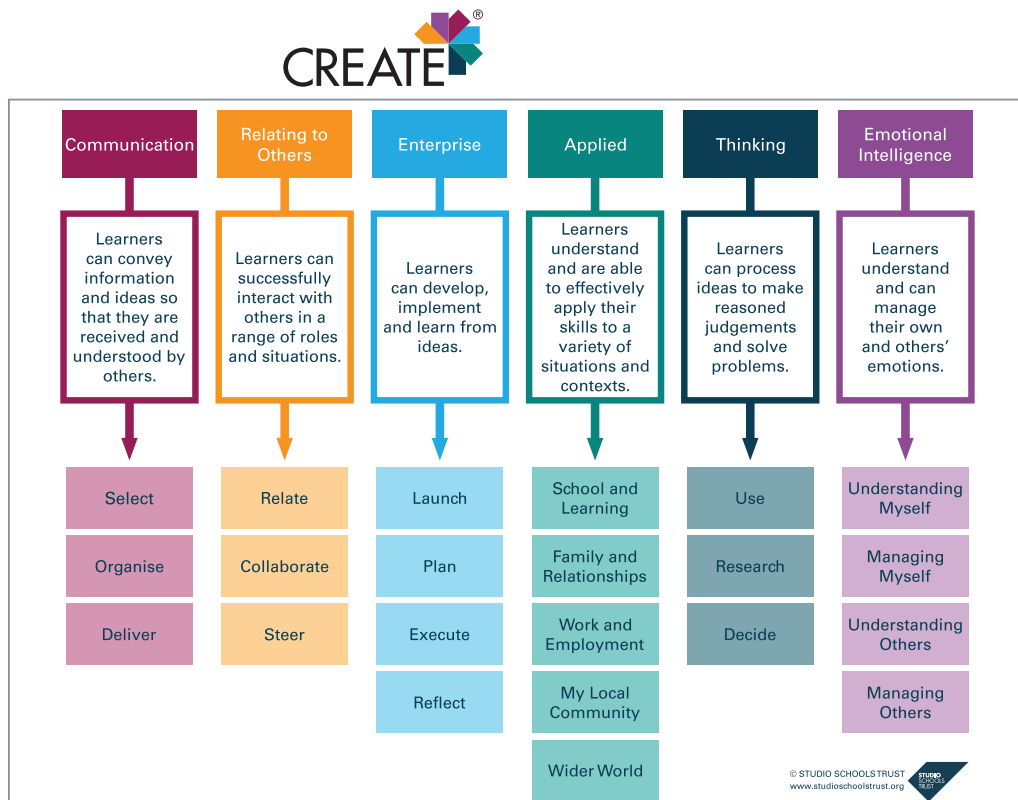


Figure 1.1: The Studio Schools CREATE®-framework of transferable skills (Source: Studio Schools Trust, 2013c)

Model components

Teaching the national curriculum, the content of the curriculum offered by Studio Schools is not fundamentally different from that of other state schools. What is unique about the Studio Schools model is the combination of the method by and setting in which the curriculum is taught, and the additional activities students and teaching staff engage in for the schools to achieve their goals. The model is based on the integration of a set of pedagogic strategies that –independent of one another– have the potential to improve student outcomes.

1. Project-based learning (PBL)
2. Personal coaching of students by a designated staff member
3. Work placements as part of students' education
4. Small school environments

1. Project-based learning

Project-based learning (sometimes called enquiry-based or problem-based learning) forms the core of the Studio Schools curriculum. Project-based learning in Studio Schools involves teaching the national curriculum principally through practical projects in the school and surrounding community, with less emphasis on teaching subject matter in the conventional classroom-based form (Studio Schools Trust, 2012a).

- What is project-based learning?

In project-based learning (PBL) students work in groups to solve challenging problems that are authentic (rooted in reality), curriculum-based, and often interdisciplinary (Solomon, 2003). It is sometimes interchangeably referred to as problem-based learning or enquiry-based learning. The approach was pioneered by Barrows (1996) who used it in order to better reflect the daily reality of medical practice in undergraduate medical education. The defining features of PBL according to this original approach are presented in Box 1.1. However, no universal definition of PBL exists, and authors conceptualise the term differently. For example, according to Ravitz (2008), PBL at least requires 1) extended student investigation, 2) in-depth inquiry into a topic, 3) some degree of student self-direction or choice and 4) presentation by students of their findings, results or conclusions. Moursund (2003) approaches the concept from a different angle and associates it with authentic content, authentic assessment, teacher facilitation but not direction, and explicit educational goals. These are both loosely defined conceptualisations of PBL, and a generally accepted statement of the defining (and distinguishing) features of PBL is lacking.

As Thomas stated in his review of early research into PBL:

‘The diversity of defining features coupled with the lack of a universally accepted model or theory of Project-Based Learning has resulted in a great variety of PBL research and development activities’ (Thomas, 2000).

1. Problems form the organizing focus of projects and stimulus of learning. An unstructured driving question of focus needs to be asked, which reflects a challenge likely to be faced in reality.
2. Problems are a vehicle for the development of problem-solving skills. Problems therefore need to be presented the same ways as they occur in the real world: only limited information is available from the outset, and students receive new information upon their own self-directed inquiries, study and research.
3. Learning is student-centered. Students self-identify what they need to learn from subject content and it is up to them to derive the key issues of the problems they face, define their knowledge gaps, and pursue and acquire the missing knowledge.
4. Teachers act as facilitators and tutors, asking students the questions they should be asking themselves in order to solve the central problem, and supporting students to take ownership of the learning. They do not lecture or provide information, unless the latter has been requested by students in order to solve the problem.

Box 1.1: Barrows’ characteristics of Project-based learning, used as a model for Studio Schools (Barrows 1996, 2002; Studio Schools Trust, 2012b)

- Rationale and theoretical base

Two fundamental principles underlie project-based learning. First, advocates of PBL recognise a need to nurture students’ ability to *transfer* the things they learn and skills they acquire in school to settings beyond school. Transfer can be conceptualized as ‘the ability to do X under conditions Z when taught to do X under conditions Y’ and forms an essential aspect of competent performance (Alexander & Murphy, 1999). However, transfer of knowledge and skills obtained in one context to other contexts requires training that is often not provided in conventional Western education systems. As a result of teaching students single subjects such as mathematics, history, or modern languages in isolation, many students continue to treat concepts or procedures in a fragmented or unduly constrained manner (Alexander & Murphy, 1999; Byrnes, 1996). At the root of this problem lies the concept of *situated cognition*, which acknowledges that learning and knowledge are situated in activity

bound to social, cultural, and physical contexts (Bereiter, 1997). If knowledge and skills are to be transferred across settings, they need to be taught in such a way that facilitates this process. Bereiter illustrates this with an interesting metaphor:

‘Run rats in a maze under typical tightly controlled conditions and they learn a fixed route to the goal. Change things a little and they are lost. But let them run around on their own and they quickly learn the whole maze, so that they can get from wherever you drop them to wherever they want to go by an efficient route.’ (Bereiter, 1997)

The idea that situated cognition –a threat to an individual’s ability to transfer knowledge across contexts– can be overcome by appropriate education is also recognised by Spiro’s *theory of cognitive flexibility* (Spiro et al., 1988). To achieve cognitive flexibility (i.e. to master complexity and the ability to transfer knowledge and skills across settings), students must attain a deeper understanding of content material, reason with it, and apply it flexibly in different contexts. Spiro and colleagues found that students often oversimplify complex material, and that this is associated with the conventional classroom-based teaching of single subject content to students. This lack of complexity in students’ reasoning and the lack of cognitive flexibility that is the result of this may be overcome by teaching subjects in a context simulated to reflect the complexity that is often encountered in reality, which is what PBL aims to establish. The second underpinning of PBL is the assumption that students may learn best when *obtaining knowledge and skills as means to an end*. Work by Schank et al. (1994) on ‘goal-based scenarios’ and by Alexander & Murphy (1999) on the conditions that contribute to successful transfer illustrates this. Schank and colleagues note that outside school, people typically learn during experiences while addressing desired goals and the learning of skills takes place within some authentic activity (i.e. an activity that reflects a realistic situation). Therefore, when students learn subject content as a means to an end, not only do they see the potential

utility of the things they learn, they also learn content in the same way they would learn it in the 'real world' (Schank et al., 1994). This facilitates the transfer of knowledge and skills acquired within the relatively structured school environment to the more ill structured situations in which it should be applied (Alexander & Murphy, 1999).

- The evidence base for project-based learning

The absence of a consistent definition of the components of PBL limits the build-up of a cumulative evidence base for claims about the effects of the approach on student outcomes. Several meta-analyses have been conducted, all of which concluded in favour of PBL versus other educational approaches for affecting a variety of student educational outcomes. However, these suffered from significant heterogeneity with regards to population (age, level of education) and intervention (actual activities taking place) across individual studies or contained no description of the actual form of teaching in individual studies (Gijpels et al., 2005; Vernon & Blake, 1993; Walker & Leary, 2009). Therefore, although PBL is a promising innovation in education, its effectiveness for educational outcomes, including engagement with education and the development of transferable skills, has not been established conclusively.

2. Personal Coaching

In Studio Schools, all students are assigned a personal coach who will meet with them on a regular basis to develop their own personalised learning plan. The aim of this component of the model is to help students navigate and personalise their curriculum, allowing them to learn in a way that best meets their interests and needs, and to track their own progress (Studio Schools Trust 2012a).

- What is personal coaching?

A variety of definitions of coaching can be found in the literature, and the term is often used interchangeably with terms such as ‘mentoring’ and ‘counselling’. Some describe the key difference between coaching and other helping roles as the coaching expectation of the achievement of specific goals and a solid commitment to planned action (Zeus & Skiffington, 2002). Others define it as ‘working with [people], with the emphasis on enhancing personal growth rather than fixing problems’ (Campbell & Gardner, 2005). Thomas and Smith recognise that coaching differs most from other supportive relationships such as mentoring and guidance in that it does not offer advice (i.e. suggestions for what to do) nor do coaches suggest that they are experts. Rather, coaching is a solutions-focused process that enables independent thinking, supports individuals in examining their own strengths, and encourages people to take responsibility for finding their own solutions to their problems or challenges (Thomas & Smith, 2009).

- Rationale and theoretical base

In the fields of education and business, coaching is primarily devoted to improving individual performance (ECI, 2004; Kilburg, 1996). Several change processes lie at the heart of such improvements. First, working with a coach, people are often better able to view their behaviour objectively than without such reflection and feedback (Ringwalt et al., 2009). Coaching provides a process of honest feedback and review that yields the benefits of an outsider’s perspective on one’s performance and behaviour (Katz & Miller, 1996). When being coached, people also learn to better recognise their strengths, which makes it easier for them to examine aspects of their work or behaviour that could be improved, and can thus prompt better self-reflection and analysis (Thomas & Smith, 2009; Veenman et al., 1999). Second, as became clear

from some of the definitions presented in the previous section, coaching is a process that aims to enable people to come up with their own solutions to problems or challenges that suit them personally, and it is rooted in the belief that people can find their own next steps. It thus avoids the potential drawbacks of more directive forms of support that may lead to solutions that may not always suit the capabilities and circumstances of the coached individual (Thomas & Smith, 2009). Finally, the process of coaching may help establish strong supportive relationships between a coach and a coached individual. This may help those that are being coached to not only actively engage with their work but may also make them feel valued, which fosters their commitment to joined goals (ECI, 2004). Such quality relationships can support people when they engage in new activities with which they might feel uncomfortable, and creates opportunities for learners to practice new skills and ideas without feeling exposed (Katz & Miller, 1996; Ringwalt et al., 2009).

- The evidence base for personal coaching

Some primary evidence indicates that coaching may have the potential to improve students' learning and wellbeing (Campbell & Gardner, 2005). However, reviews of the academic literature have shown that very little empirical evidence exists on the effectiveness of coaching and factors that might mediate or contribute to such effects (Feldman & Lankau, 2005; Griffiths, 2005; Kilburg, 1996). This results in the evidence base for personal coaching being in an inconclusive state, and the strong support it receives in business and education is often based on the surveyed opinions of those involved in the process (for example, see ECI, 2004; Thomas & Smith, 2009).

3. Work placements

In Studio Schools, all students spend a significant portion of their weekly time in work placements, in which employers are engaged. Work placements are intended to be as realistic as possible, including actual settings of employment and payment for students' work. Students in years 12 and 13 (age 16-18) are expected to sign a contract of employment with a local employer, and work for at least 2 days per week. Students in years 10 and 11 (age 13-15) are expected to participate in a minimum of 4 hours of (unpaid) work per week (Studio Schools Trust, 2012a). This model component has a dual aim. First it is assumed that real work placement will increase students' engagement with school as it directly shows the benefits of what is learned in practice. Second, real work placements aim to better prepare students for the labour market than schooling without such placements. In Studio Schools, work placements are intended to reinforce non-cognitive skills obtained through project-based learning and vice-versa (Studio Schools Trust, 2013b).

- Rationale and theoretical base

The idea that employment as part of schooling –often referred to as work-based learning- can improve important student outcomes is not a novelty. First and foremost, realistic work experience is seen as a means to prepare students for employment upon finishing their education. In the 1990s, the observation that youth were insufficiently prepared by secondary schooling for a labour market subject to continuous change was of great concern to governments across the world, as many youth did neither enrols into higher education nor into employment after graduating secondary school (Bailey & Hughes, 1999; Ryan, 2001). As a response to this situation, studies and reports arose on students in work-based learning programs in (amongst others) the UK, the US and Australia (Fullarton, 1999; Hamilton &

Hamilton, 1997; Howieson, 1990). In those countries, numbers of students in work-based placements at that time were as high as up to 65 percent in the 1990s. Positive results for students' employability as a result of gaining work experience may come through several routes. One of these is the validation of previous work experience by employers when hiring (Hollenbeck, 1995). As a result of gaining work experience, students can also develop a (better) understanding of the everyday functioning of industries (Hershey et al., 1999), which may help them obtain a (better) understanding of their desired career paths and goals (Stasz & Brewer, 1999). Second, there is some indication that employment as part of schooling may indeed increase students' engagement with their education. As real work experience as part of the curriculum gives students the chance to complement classroom-based education with the practical setting of a workplace, it may be a strategy to improve students' interest in and engagement with their education (Kemple & Snipes, 2000).

- The evidence base for work placements as part of secondary education

The primary argument for real work experience as part of the Studio School curriculum is that it increases student's employability upon graduation. Some empirical reports support this notion. Some non-randomised studies focusing on post-graduation employment as an outcome report that students who were employed as part of their curriculum were more likely to be employed within a year after graduation from a secondary education institution (e.g. Arum & Shavit, 1995). A secondary argument for real work experience as part of secondary education is that it can increase student's engagement with their education. Few studies seem to have empirically tested this assumption. Using a randomised controlled design to evaluate the effects of US Career Academies, Kemple and Snipes (2000) found that the approach, of which career oriented work-based learning forms an integrated part, had

a positive effect of student's engagement with education. The evaluation showed that the Career Academy approach improved student attendance and decreased dropout rates. This was particularly the case amongst students at high risk of dropping out. However, the authors point out that, some students in Career Academies in which perceived peer and teacher support was relatively low were slightly more likely to drop out and had slightly lower rates of attendance than control group students from schools subject to the same negative perceptions. This suggests that the work experience component of the Career Academies model may not be enough as a stand alone component to have a positive effect on drop out and attendance rates, and that other (intentional or unintentional) aspects of the model may have a substantial influence on its effectiveness. Evidence regarding the effects of gaining work experience on students' academic performance and opportunities is equally scarce. This is likely due to a variety of definitions of work-based learning, and of the quality of students' work placements. As Hughes and colleagues conclude after their review and primary studies on work-based learning and how it affects students' academic skills:

'Our experience with work- based learning teaches us that one cannot easily generalize about its impact' (Hughes et al., 1999)

Primary studies on the topic report different results. For example, whereas Arum & Shavit (1995) found that work-based learning or a focus on vocational education may have a negative impact on students' enrolment into higher education, evidence from US career academies shows that this does not need to be the case (Kemple & Snipes, 2000). Hughes et al. (1999) argue that this may heavily depend on the quality of work placements. Different study designs (experimental vs. quasi-experimental) may also be an important factor for diverging results, because students who self-select into

work-based learning or other vocational programs are likely to differ substantially from students who opt for mainstream education. This poses a particularly large risk of selection bias to non-randomised studies in this field.

4. Small school environments

Studio Schools are small schools of around 300 students, offering a personalised and supportive learning environment for their students (Studio Schools Trust, 2012a). On the one hand, this directly targets students' attainment in and engagement with school. On the other hand, a small school facilitates the other components of the Studio Schools model. Overall, small schools should lead to clarity of individual students' needs to staff and that no child is left behind (Studio Schools Trust, 2012a).

- Rationale and theoretical base

The sheer size of a school in itself is unlikely to explain any positive or negative outcomes for its students. It is mostly the interaction between students and between them, teachers and parents that might be better facilitated by small school environments (Newman et al., 2006). Such interaction is fostered by the more personalised learning environment that schools that are small in size can offer versus larger schools (Darling-Hammond et al., 2002), teachers and students are likely to know one another better than in the former. This implies that a reduction in size is not enough for a school to succeed. Schools should not only be small in size but also in function (Bloom et al., 2010), meaning that the school organisation (reduced teacher load, meetings between teachers to discuss students' progress and problems) should take advantage of the opportunities provided by a reduced student body size. Wasley & Lear (2001) argue that if small schools continue to organise themselves the way large comprehensive schools do, with a bureaucratic leadership structure and

prescribed (contained) teacher duties, most of the benefits of small schools disappear as they become ‘small big schools’.

- The evidence base for small school environments

The evidence regarding the effects of small school environments on students’ attainment in and engagement with school is mixed. Some primary studies report positive effects of small school environments on (among other outcomes) students’ educational attainment in the form of standardised tests (Raywid, 1998), engagement in the form of attendance, dropout rates and years of completed education (Fine, 1994; Oxley, 1995), and less violence in schools (Franklin & Crone, 1992). The positive results for students of small schools may also lead to increased cost-effectiveness of such schools versus larger secondary schools on a per graduate basis (Steifel et al., 1998). One factor that inhibits a reliable estimate of the effects of small school environments on student outcomes is that the majority of studies on the topic use suboptimal designs. However, even studies with high quality designs offer inconclusive evidence. For example, although an experimental study conducted in New York City secondary schools showed that students in small schools were more likely to graduate and to do so timely, the size of the school was only one of the components of a larger educational reform programme (Bloom et al., 2010), thus limiting unconfounded causal inference. Additionally, a systematic review by Newman and colleagues (2006) concluded that the effects of small schools on student behaviour are equivocal, and that –although some negative effects were found on attainment outcomes for a decreasing school size– no valid conclusions could be drawn for student attainment. However, the review did conclude that teachers and students at smaller schools were more likely to have positive perceptions of their school environments.

1.3.3 Evaluating Studio Schools

Studio Schools are spreading steadily across England, and affect the lives of an increasing number of students. The potential impact of this educational innovation is vast, given the increasing momentum the Studio Schools movement is gaining. Intervening with people's lives brings along the responsibility to do more good than harm, and to evaluate practices in order to know if one is doing so.

The limited evidence base for Studio Schools

Studio Schools can, from an implementation perspective, best be described as a *community centred model* (i.e. one that started with a community and its needs rather than with researchers and their research) (Wandersman et al., 2003). This implies that they are not developed from a perspective of evaluation science, but rather pragmatically and on a basis of pedagogical experience as well as innovative views on education. Studio Schools claim to be using a model of tried and tested components (project-based learning, personal coaching, real work experience, small schools) integrated into one approach. As the Studio Schools Trust formulates it:

“What makes Studio Schools unique is that they bring these [components] together, for the first time, into one whole school model” (Studio Schools Trust, 2012a).

Although all of the individual components of the Studio Schools model have indeed been ‘tried and tested’, the previous section has shown that no conclusive evidence exists for any of these components when implemented as stand alone interventions or programs. Therefore, although the theoretical underpinnings of these components appear promising, there is no empirical basis that provides conclusive support for their effectiveness. Additionally, the assumption that the integration of promising (or even effective) programs or services will lead to mutually reinforcing synergy and as a result better outcomes does not uphold automatically. When interventions are

implemented differently from how they were implemented when they were evaluated, or when additional interventions are offered alongside, we cannot be certain that the effects of (an) original evaluation(s) will occur (Lane et al., 2004; Mihalic, 2004). Thus, to support any claims regarding the potential of Studio Schools to reach their targets with empirical evidence, an evaluation of the model is needed.

Challenges to evaluating the Studio Schools model

Ultimately, only an outcome evaluation can determine if Studio Schools are an effective school model to increase students' engagement with school and develop their transferable skills. However, and central to this DPhil project, the current state of development of the Studio Schools movement and model pose some challenges that undermine the usefulness of the results of a trial if were to be undertaken at this point.

The lack of an explicit and coherent programme theory

It can be argued that the primary purpose of evaluating an intervention or program model is to determine whether it is effective in reaching its objective. However, in order to make full use of an intervention's evaluation, we also need to know how and why interventions work (Weiss, 2004). If we do not understand the mechanisms by which interventions are (in)effective, outcome evaluations cannot guide policy makers, practitioners and other stakeholders regarding how to intervene successfully with social issues. Central to understanding the mechanisms of effect of interventions is the specification of a programme theory or *theory of change*. Theories of change describe a causal chain of activities intended to produce a positive intervention outcome (Fraser et al., 2009), and may also include predictions of how this mechanism is affected by contextual factors (Pawson & Tilley, 1997). Specifying the theories of change of interventions in detail and measuring to what extent theorised

relationships uphold is essential in order to determine why and how interventions may work (Weiss, 1997). Additionally, the accumulation of evidence from theory-based evaluations (those that evaluate theoretical underpinnings alongside effects) is important to develop effective theoretical models on which new social interventions can be based (Chen & Rossi, 1987). Although the objectives and core model components of Studio Schools have been defined, this model lacks a coherent and explicit theory of change. No specification has been given of how the mechanisms by which the model components might lead to the desired outcomes, and how these components might interact with one another in this process. Connell & Kubish (1998) defined three quality criteria for theories of change:

1. A theory of change should be plausible. Do evidence and common sense suggest that the activities, if implemented, will lead to desired outcomes?
2. It should also be doable. Will the economic, technical, political, institutional, and human resources be available to carry out the initiative?
3. Finally, it should be testable. Is the theory of change specific and complete enough for an evaluator to track its progress in credible and useful ways?

The absence of an explicitly defined theory of change makes it impossible to assess its quality, which inhibits any theoretically valid presumptions about its plausibility, doability, or testability. This would make it impossible to incorporate a theory-based process evaluation into an outcome evaluation of the Studio Schools model at this point, which hinders the explanatory power of an outcome evaluation.

Heterogeneity of implementation

Studio Schools are founded and operated locally, and no explicit prescription (e.g., a manual, prescribed curriculum) exists that specifies how schools should work with the model or how it should lead to desired outcomes. With little guidance offered to schools with regards to how the model components should be implemented in order to work effectively, there is a realistic chance that implementation varies extensively

across schools. For example, preliminary interviews with the Studio Schools Trust confirm that the extent to which curriculum content is taught to interdisciplinary projects as opposed to single subject teaching is likely to vary greatly between schools. This poses a serious threat to the validity of an outcome evaluation of the Studio Schools approach. If implementation of a model varies to the extent that sites are not implementing the same coherent set of theoretical principles, an outcome evaluation would not actually test the effectiveness of the intended model and would therefore miss its primary target.

Preparing Studio Schools for outcome evaluation

At this point it is unclear how the components of the Studio Schools model should lead to the desired targets of young people's increased engagement with education and their development of transferable employability skills. Furthermore, because the Studio School model is not specified in detail and implementation is led locally, it is unclear whether schools are in fact implementing the same model. As a result of this, an outcome evaluation of Studio Schools conducted at this point would not be able to explain the mechanisms for the effects (or lack thereof) it would find. Such an evaluation would also be unable to lead to an efficiently scale-able program. In addition to this, Studio Schools are currently insufficiently oversubscribed (i.e. more applicants apply than there are spaces available) to allow for a feasible pragmatic randomised controlled trial based on an admissions lottery. Therefore, much can be gained at this point from formative research that focuses on the process of Studio Schools' functioning. In order to prepare the Studio Schools approach for a valid outcome evaluation in the form of a RCT, two steps need to be undertaken.

1. Elucidate the Studio Schools theory of change

Even when the theory of change or the structure of a program is not specified explicitly, it is often present implicitly (Chen, 1990). However, to systematically guide a process- or outcome evaluation and program implementation, an explicit theory of change is required (Connell & Kubisch, 1998; Eldridge et al., 2005; Weiss, 1995). For this purpose, the specification of program logic models (a graphic depiction of a theory of change) has been recommended as they are easy to comprehend by a variety of stakeholders (Cooksy et al., 2001; Trevisan, 2007). Such models can be developed with the help of purveyor organisations even when not specified explicitly beforehand (Oosthuizen & Louw, 2013). The elucidation of an explicit theory of change is particularly relevant for Studio Schools as the implementation of this theory is to be led locally.

2. Assess implementation practices

Once an explicit theory of change has been specified for the Studio School model, the extent to which different Studio Schools are actually implementing the same model needs to be determined. Although implementation tends to vary across settings when implementing social interventions, the need to implement programs with some fidelity to their intended models is vital for comparing like with like and understanding their effects when evaluating them (Durlak & DuPre, 2008; Mihalic, 2004). This is of particular importance for the Studio School model, as local schools are explicitly encouraged to tailor the model to the needs of their communities and to lead implementation locally (Studio Schools Trust 2012a). The importance of monitoring implementation and how local implementation and adaptation affect implementation fidelity are further discussed in part 1.3 of this chapter. Furthermore, the different ways by which implementation fidelity might be conceptualised and how this affects

the evaluation of universally offered but locally tailored intervention models will be discussed.

1.3.4 Section summary

Studio Schools are a new type of state secondary school model that aims to provide an alternative to conventional secondary schools in England. The Studio School model has been developed as a response to the disengagement of a large group of young people with their secondary education and the lack of focus on transferable employability skills of conventional secondary schools. This model incorporates four pedagogic strategies to improve these outcomes for their students: 1) project-based learning, 2) personal coaching, 3) work placements, and 4) small school environments. Although the academic literature demonstrates the promise of each of these strategies independently, the evidence base supporting their effectiveness is limited and their joint effectiveness cannot be assumed. The effects of the Studio School model on its targeted outcomes should be determined by a rigorous outcome evaluation. The current lack of clarity around the model and its implementation would inhibit an outcome evaluation to present meaningful results if undertaken at this point, because it would be unclear to what factors any effects (if found) would be attributable. Before a meaningful outcome evaluation of Studio Schools can be undertaken, these issues –the lacking of an explicit theory of change and a lack of insight in implementation practices– need to be resolved.

1.4 Evaluating the implementation and adaptation of interventions

1.4.1 Why evaluate implementation?

The monitoring of program implementation has been described as an essential feature of effectiveness trials and dissemination efforts of interventions (Durlak & DuPre, 2008; Dusenbury et al., 2005). The relevance of implementation science for the field of intervention research manifests itself across three main dimensions: 1) it aids in understanding and interpreting the effects of interventions, 2) it facilitates the accumulation and synthesis of research evidence, and 3) it is necessary for the successful dissemination of evidence-based practices.

Understanding the effects of interventions

Using rigorous evaluative research designs to evaluate the effects of interventions on outcomes lies at the heart of evidence-based practice. Even when tested through randomised controlled trials, however, no confident connections can be made between programs and outcomes without knowing to what extent the program under test was actually implemented (Durlak, 1998; Perepletchikova & Kazdin, 2005). For this purpose, process evaluations are often integrated into experimental designs (Bonell et al., 2012; Oakley et al., 2006). To enable such evaluations to properly test the theoretical assumptions underlying interventions, it is essential that implementation put these assumptions into practice across testing sites. Without evaluating what is put into practice, trials are at risk of falsely concluding that services are ineffective when in reality implementation is so poor that it leads to misinterpreted outcome analyses. This error is known as type III error (Dobson & Cook, 1980; Domitrovich & Greenberg, 2000). The difficulty to deliver treatment variables in a uniform way is seen by some as the chief challenge to evaluating behavioural (social) interventions

using randomised controlled trials (Carroll et al., 2000). As a result of this, the monitoring of implementation as part of primary effectiveness studies is essential to understand how and why interventions show (or fail to show) effects.

Accumulation and synthesis of research evidence

Apart from rigorous primary studies, an essential part of evidence-based practice is the combining (synthesis) of the best available primary evidence for the effects of interventions with high quality systematic reviews (Chalmers & Altman, 1995). Systematic reviews facilitate the accumulation of an evidence base for interventions, and may be at the root of social change through policy and practice by providing directive guidance on the effectiveness of a given intervention for a given problem. For systematic reviews –particularly those that include meta-analyses– to provide robust and meaningful conclusions about interventions’ effects, a good understanding of how interventions that were evaluated in included primary studies were implemented is essential. Central to this is the concept of *homogeneity*: the extent to which included studies have evaluated the same intervention. An understanding of how implementation practices affect the (statistical and clinical) homo- or heterogeneity in meta-analyses is necessary to provide adequate guidance to practitioners, policy makers and other stakeholders (Montgomery et al., 2013). Indeed, examples exist of systematic reviews losing much of their conclusive value because of very large variance in implementation (Littell et al., 2005; Wilson et al., 2003). There is much debate in the literature about what it is that needs to be implemented homogeneously. For example, whereas some authors argue for the strict adherence to detailed implementation manuals, others hold the view that what matters is fidelity to the core set of underlying theoretical principles of programmes (Hawe et al., 2004a; Mihalic, 2004). This debate will be further discussed in paragraph 1.3.4.

Dissemination of evidence-based practices

For evidence-based interventions to fulfil their promise when becoming integrated into common service or policy, the change mechanisms that made original models effective need to be preserved. Because the effects of evidence-based services are related to their implementation, this causes some to argue that they should be implemented with fidelity to the original (tested) model (Arthur & Blitz, 2000). Indeed, a lack of implementation fidelity may for a large part cause the failure of interventions that have been proven effective in controlled settings to bring about the same effects in real life contexts (Bellg et al., 2004). For example, the evaluations of the Botvin Life Skills Training Program in New York urban schools showed that significant effects on students' smoking, alcohol, and marijuana use were only found when the program was implemented with fidelity (Botvin et al., 1990). For dissemination of interventions on the large scale to be effective, researchers need to understand the processes required to implement the intervention consistently and with high quality (Breitenstein, 2010). Indeed, high quality implementation has been shown to be more likely when core intervention ingredients are defined in advance and then systematically monitored to ensure compliance (Gresham et al., 1993).

1.4.2 The importance of implementation fidelity

The concept of implementation fidelity is essential to evaluating the implementation of social interventions, and is often held to be in conflict with the need to adapt interventions to different contexts. Below, implementation fidelity will be conceptualised, after which its relevance for intervention research will be addressed. This will be followed by a discussion of attempts by researchers to secure fidelity of implementation as part of evaluation and dissemination efforts. Adaptation of

interventions will subsequently be discussed in section 1.3.3 and the fidelity-adaptation tension in section 1.3.4.

What do we mean by implementation fidelity?

Implementation fidelity (also referred to as programme implementation, programme integrity or treatment fidelity) refers to the extent to which a program or intervention is implemented as intended. This definition leaves room for multiple interpretations. Some authors define fidelity narrowly as intervention providers' compliance with treatment protocols (Breitenstein et al., 2010; Mowbray et al., 2003), whereas others use definitions that incorporate multiple facets of implementation. In their review of the effects of programme integrity on outcomes, Dane & Schneider (1998) recognised five dimensions of implementation fidelity: *adherence, exposure, quality of programme delivery, programme differentiation, and participant responsiveness*. Other notable authors have also used these dimensions for their definitions of fidelity or integrity (Domitrovich & Greenberg, 2000; Mihalic, 2004). Instruments developed for the purpose of evaluating implementation across multiple studies, such as the Yale Adherence and Competence Scale (Carroll et al., 2000) and the Oxford Implementation Index (Montgomery et al., 2013), also focus on the delivery of intervention components and the uptake of those by participants. Although multiple definitions of implementation fidelity exist, authors generally stress the importance of implementing the 'core components' (sometimes called active components, active ingredients, or essential components) of an intervention. These are the components primarily responsible for an intervention's outcomes (e.g., Durlak, 1998; Elliott & Mihalic, 2004). It needs to be noted here that the fact that the core components of interventions are often unknown or insufficiently specified causes some authors to

argue that implementers should strive for 100% fidelity to what was tested in trials (Mihalic, 2004).

How implementation fidelity affects intervention research

The importance of adhering to original intentions across sites when evaluating and implementing interventions is widely regarded as essential to success and the extent to which this is done is increasingly assessed through process evaluations integrated into experimental evaluations (Bonell et al., 2012; Oakley et al., 2006). Firstly, fidelity influences the validity of claims about interventions' effectiveness. Even when tested rigorously through randomised controlled trials, it is unclear what caused the connection between a program and observed (lack of) effects. For example, when a Cochrane review concluded that Multi-Systemic Therapy was no more or less effective than no treatment for treating youth social, emotional and behavioural problems (Littell et al., 2005), the developers of MST claimed that this conclusion relied heavily on the inclusion of one particular study in which MST was implemented with questionable fidelity (Henngeler et al., 2006). The claim that the largest study (which carried the most weight in the meta-analysis of the review) suffered from major issues of fidelity was confirmed by the authors of that study. The result is an *inconclusive* state of the evidence around a well-researched intervention. Secondly, once a convincing evidence base supports an intervention, implementation fidelity is considered a key ingredient for an effective replication of such interventions. For dissemination of interventions on the large scale to be effective, researchers need to understand the processes required to implement the intervention consistently and with high quality (Breitenstein, 2010) across different settings. Because of this, some argue that for interventions to be effective on a society-wide

level, it is equally important they be implemented with fidelity as for them to be rooted in a sound theoretical and evidence base (Webster-Stratton, 2004).

Attempts to maintain implementation fidelity

A couple of strategies are used to maintain implementation fidelity as part of evaluation or dissemination efforts. First, fidelity is often monitored during the evaluation of interventions. This ensures that the assumed relationship between implementation and outcomes can be tested, and lessons can be learned as to how to improve implementation practices (Elliott & Mihalic, 2004; Mowbray et al., 2003). This is done by specifying and operationalising intervention components *a priori*, and ideally –when resources allow– using multiple methods (including quantitative and qualitative reports from implementers and participants and on-site observation) to measure the extent to which they are implemented as intended (Dusenbury et al., 2005; Gresham et al., 1993). Fidelity is also monitored frequently during (early) efforts to disseminate evidence-based practices, in order to ensure that implementation practices in uncontrolled settings match those that were tested in controlled trials (Mihalic, 2004). Second, the content and dose of interventions components are often manualised in order to achieve a high level of fidelity when delivered by others than those who designed them. Successfully manualised evidence-based interventions include parenting programs, substance abuse prevention programs, and juvenile offence prevention programs (Botvin & Griffin, 2004; Henngeler et al., 2006; Sanders, 1999; Webster-Stratton, 2004). The main purpose of intervention manuals is to ensure that the different core techniques or activities are standardised and implemented the same way in real world settings as when designed and, in some occasions, tested in controlled trials. Third, researchers have developed frameworks to provide practitioners with guidance to the implementation process.

As part of these strategies, implementation fidelity is typically conceptualised as the degree to which standardised intervention activities or techniques that have been derived from programme theory by intervention developers, are implemented (for example, see: Borrelli et al., 2005; Dufrene et al., 2005; Lane et al., 2004, Hardeman et al., 2008).

1.4.3 Adaptation of interventions

The adaptation of interventions by those who implement them is a concept at the centre of the study of implementation. A definition of adaptation and discussion of its relevance for intervention research will be provided below, after which reasons for the adaptation of interventions will be addressed. Following this, some ways by which interventions might be adapted will be discussed.

What do we mean by adaptation?

When social interventions are implemented under natural conditions, they are rarely delivered in the same way they were originally designed and evaluated but often adapted to fit different contexts (Dariotis et al., 2008; Dusenbury et al., 2005; Gottfredson & Gottfredson, 2002). Adaptations to programmes may be purposefully planned prior to implementation or made spontaneously without prior planning in attempts to make them more suitable or appealing in specific contexts or for specific participants (Lee et al., 2008; Moore et al., 2013). One factor that may prompt such efforts is the need for ensuring cultural appropriateness: in order to match different ethnic or cultural populations or contexts, social interventions sometimes require *cultural adaptation* (Castro et al., 2004; Kumpfer et al., 2002; Ringwalt et al., 2004). Some authors recognise a more general need for *local* or *contextual adaptation*: social interventions require adaptation for and by local stakeholders to facilitate uptake and

effectiveness, which may be caused by a range of contextual factors (Lee et al., 2008; Ozer et al., 2010). Adaptation of interventions may also occur unintended as a result of features of implementation systems such as available resources and providers' required skills (Durlak & DuPre, 2008; Ringwalt et al., 2004). Whether planned or unplanned, intended or unintended, the term adaptation implies that something is changed in an original intervention model or design. Defined by its implications for implementation science, adaptation of interventions can therefore be conceptualised as the *addition, subtraction, or modification* to an original model, the quality of delivery of that model, or participants' responsiveness to implementation of that model (Moore et al., 2013).

How adaptation affects intervention research

Adaptation of tried and tested social interventions when implemented in uncontrolled settings has been described as the 'rule rather than the exception' (Moore et al., 2013; Schinke et al., 2002), and it is widely recognised that in order to fit the needs of different communities in different contexts, adaptation of innovations is bound to occur (e.g. Berwick, 2003; Dane & Schneider 1998; Domitrovich & Greenberg, 2000; Rogers, 2003). As evidence-based intervention models are adapted, the validity of their underlying evidence base becomes questionable, as what is implemented may no longer reflect what was tested. At the same time, unless adaptations are carefully monitored and tested for their effects, it is unclear to what extent these might be beneficial or harmful to those receiving an intervention (Castro et al., 2004). For these reasons, an understanding of the adaptations that are made to social interventions in real world settings, and why and how such adaptations are made is important in order to understand their effects and to guide successful dissemination efforts.

Why might social interventions be adapted?

The examples outlined in section 1.4.2 illustrate that in some cases it is possible to successfully standardise intervention techniques or activities for implementation across different settings. Such interventions are highly structured, operate within controllable contexts, and target behaviours or factors that are quite universal (such as basic parenting skills), which may for a large part explain their success. In many cases however, the strong focus on standardising activity content and delivery that is often at the core of conceptualising interventions may be at odds with the flexibility demanded of these by the diverse social settings in which they are to be implemented. This may particularly be the case for *complex social interventions* that are designed to be implemented on a large scale. Although it is difficult to define a complex intervention distinctively, they possess five defining features according to the MRC (Craig et al., 2008):

- The number of interacting components within the experimental and control interventions
- The number and difficulty of behaviours required by those delivering or receiving the intervention
- The number of groups or organisational levels targeted by the intervention
- The number and variability of outcomes
- The degree of flexibility or tailoring of the intervention permitted

The greater the complexity of an intervention's programme theory (especially the number of different interacting activities that need to take place), and the greater the complexity and specificity of the social system with which it interacts, the more likely the form of its components might be adapted across different settings. Research demonstrates that in some cases adaptations might indeed be necessary to facilitate optimal intervention effects (Durlak & DuPre, 2008). In other cases, adaptation of intervention components is not strictly necessary to secure or enhance effects but nevertheless preferred by providers, participants, or other stakeholders (Ozer et al.,

2010). It is therefore not surprising that implementation fidelity to original models never reaches 100% (Durlak & DuPre, 2008). Some go as far as arguing that adaptation is the defining characteristic of human production (Pagel, 2012). Multiple factors may lie at the cause of this. First, specific features of the implementation system and social context may prompt adaptations to intervention models. For interventions to be successfully disseminated they must be feasible and acceptable, and achieve adequate coverage across settings (Bonell et al., 2006). Some authors mention the importance of organisational capacity and (a shortage of) available resources; it may for example be difficult to find adequate staff or to recruit the right participants (Durlak & DuPre, 2008; Moore et al., 2013; Ringwalt et al., 2004). Others recognise characteristics of implementers or providers as a cause of adaptation (Hill et al., 2007; Machin & Fogarty, 2003). Such characteristics include providers' self-efficacy, skills proficiency, and attitudes to an intervention or its theoretical underpinnings (Durlak & DuPre, 2008; Ringwalt et al., 2004). Second, interventions may be adapted to better match participants' characteristics in the eyes of providers or communities. Typically, providers intend to serve a population and are not primarily interested in testing a theoretical model (Backer, 2001). Therefore, they often wish to adapt programmes to directly address the needs or characteristics of their clients (Lee et al., 2008). Differences in prevalence and patterns of manifestation of a social problem may exist across different populations, as well as difference in prevalence and patterns of risk factors and predictors for that problem may exist across different populations (Resnicow et al., 2000). Cultural differences between target populations and tested populations can also prompt adaptation of services (Moore et al., 2013). Often, communities are actively engaged in the contextual adaptation of interventions to suit their specific circumstances or needs. This so-called 'community buy-in' is

considered by some to be essential in order to sustain local participation in interventions (Castro et al., 2004). This may be one of the reasons why granting agencies increasingly demand that intervention approaches be relevant to the specific needs of community members and incorporate their norms into scientifically valid approaches (National Institutes of Health, 2008).

How are social interventions adapted?

As defined at the onset of this section, adaptation can occur related to every aspect of intervention implementation including fidelity, dosage, quality of delivery, participant responsiveness, and program reach (Dane & Schneider, 1998; Durlak & DuPre, 2008). Adaptations can be made to intervention procedures (e.g., time, location, process of recruitment), dosage (e.g., number of sessions, hours of intervention), content (adding or removing program components), participants (social, ethnic, or cultural groups), and cultural relevance (Bernal & Saez-Santiago, 2006; Moore et al., 2013). For example, Ozer and colleagues (2010) found that 45% of teachers they interviewed regarding the implementation of substance use and violence prevention programs made changes in the instructional format (e.g. restructuring individual activities to group activities) of these programmes and 21% of teachers incorporated new content based on real-life experience into the curricula. Kumpfer et al. (2002) describe efforts such as using culturally relevant examples and relocating the intervention setting (in this case African-American churches) in order to improve uptake of a family-based prevention program.

Although many (often simultaneously occurring) ways of adapting social interventions exist, an important conceptual distinction can be made between adaptations made to the surface structure and to the deep structure of interventions (Resnicow, 2000). The *surface structure* of interventions involves the ‘packaging’ of

intervention materials and messages in such a way that these match the preferences (i.e. language, brands names, role models) of a target population. An example of adapting the surface structure of an intervention would be the translation of English program materials into Spanish for the purpose of expanding reach to Hispanic populations in the United States. The *deep structure* of interventions reflects the mapping of intervention components to the cultural, social, psychological, ecological, and historical factors affecting the behaviours that are targeted. The deep structure of an intervention might be adapted by targeting additional or to some extent different risk factors (e.g., school drop out and family stability in addition to or as opposed to peer pressure) when delivering a substance use prevention programme to different ethnic groups (Resnicow 2000). The distinction between the surface and deep structures of interventions is conceptually related to the distinction between ‘core’ or ‘essential’ versus peripheral components (e.g. Durlak, 1998; Ozer et al. 2010; Webster-Stratton, 2004). Core components (those primarily responsible for the outcomes of interventions; explicitly captured in a programme theory) usually reflect the deep structure of an intervention, whereas peripheral components (those only marginally responsible for outcomes, and often primarily responsible for other purposes such as reach and uptake) usually reflect its surface structure. Whereas adaptations can be made to both structures of interventions, the implications for intervention research and evidence-based practice such adaptations may have are different for each structure. However, little is known with regards to these implications as it is often unclear what the core components and deep- and surface structures of interventions are (see observation cited from Mihalic, 2004 in section 1.3.2) and more empirical work on this is needed.

1.4.4 The tension between implementation fidelity and contextual adaptation

The need for interventions to be implemented with fidelity and for their adaptation to fit the different contexts in which they are to be implemented has become clear from the previous sections. However, when required simultaneously (which is often the case), implementation fidelity and contextual adaptation may conflict. This tension has been recognised for more than a decade with terms including the ‘fidelity-adaptation debate’ and ‘finding the balance’ between fidelity and adaptation (Dane & Schneider, 1998; Backer, 2001; McHugh et al., 2009). The nature of this conflict is outlined below.

How and why do the needs for implementation fidelity and contextual adaptation conflict?

As noted earlier in this chapter, the widely recognised need for implementing interventions with fidelity to their underlying theoretic models has led to a tendency to standardise and manualise the set of activities that the intervention comprises. Such activities are commonly referred to as the interventions’ core components or active ingredients (Dane & Schneider, 1998; Mihalic, 2004). In evaluative studies, implementation fidelity is monitored as adherence and response to such activities by implementers and participants (Gresham et al., 2003; Mihalic, 2004). It is generally assumed that, in the case of effective interventions, this will lead to better outcomes. This assumption has been empirically verified in some primary studies and syntheses of multiple studies although in none of these studies fidelity reached 100% (Botvin et al., 1990; Durlak & DuPre, 2008; Lipsey, 1999). Additionally, because allocation to sites that implement practices with weak versus strong fidelity is never random, it is difficult to obtain an estimate of the effect of fidelity on outcomes that is orthogonal to related influences such as provider skill and contextual factors. Nevertheless, some

researchers see adaptation of intervention activities as a potential threat or even a barrier to implementation fidelity (Breitenstein et al., 2010; Mihalic, 2004), which may result in iatrogenic intervention effects (Dishion & McCord, 1999).

However, others view adaptation as an important feature of intervention implementation and dissemination in its own right and argue that without such adaptations, both an intervention's chances of widespread dissemination and its effectiveness across settings are put at risk. Ensuring strict standardisation of the activities of complex large-scale interventions leaves little room for adopters (such as local authorities and communities) to adapt the intervention to their needs and may thus inhibit the intervention's adoption and (successful) dissemination (Castro et al., 2004; Kumpfer et al., 2002; McHugh et al., 2009). Indeed, apart from a handful of examples of well-researched and successfully disseminated behavioural change interventions, there is a widespread gap between acquired research evidence and common practice in fields such as public health, social work and clinical psychology (Glasgow et al., 2004; Mullen et al., 2008; Stewart & Chambless, 2007). Whereas there is a significant body of literature containing suggestions on how to train, motivate and monitor potential deliverers of standardised interventions (see Fagan et al., 2008; Hahn et al., 2002; Webster-Stratton, 2004), the sustainability and efficacy of such activities over the longer term can be questioned. Even if disseminated on a large scale, the standardisation of an intervention's activities may jeopardise its effects if there is poor interaction of (some) components with the specific contexts in which they are implemented. Interventions are implemented within existing complex systems, and may wash out over time if the dynamic properties of these systems are not harnessed appropriately (Hawe et al., 2009; Ogilvie et al., 2011). The simultaneously occurring needs to implement interventions with fidelity across

contexts and to adapt them to fit these different contexts therefore poses a paradox: in order to develop an evidence base for interventions, their components are standardised, while in order to disseminate those interventions their components require adaptation.

1.4.5 Recent developments in the fidelity versus adaptation debate

Novel approaches in thinking about implementation fidelity

The previous section has shown that, although practiced frequently and often sensibly, there may be risks involved in standardising the components of complex interventions, affecting their dissemination as well as their effects. For an evidence-based intervention to be replicable across settings however, some standardisation of what constitutes the intervention is necessary and the above does not imply that no standardisation of any kind is possible. Emphasis is being placed within many areas of the literature on the role of interventions' underlying theoretical principles in the standardisation of their components, and authors acknowledge that what is truly at the heart of intervention implementation are its *functions*, or the theorised change mechanisms that it should trigger (Hawe et al., 2004a). The use of logic models, graphic and comprehensible depictions of a theory of change, can assist in developing and disseminating interventions' detailed theories of change (Backer, 2001; Cooksy et al., 2001; Eldridge et al., 2005), even when working with provider organisations that have not previously developed an explicit theory of change (Oosthuizen & Louw, 2013). Evaluation of the implementation of interventions' functions could lead to evidence-based theories that can be used to design local intervention activities. The accumulation of evidence for and against theories instead of specific 'packages' of interventions has been proposed before (Ogilvie et al., 2011; Eldridge et al., 2005). Once evaluated, interventions would be manualised in line with this; clearly

specifying interventions' intended functions rather than prescribed activities. It would then be up to those using these programme theories to develop the intervention components, to involve stakeholders in moving from theory to intervention and pilot these interventions to establish if they are locally feasible and acceptable (Bonell et al., 2012), which would facilitate dissemination and sustainability (Rogers, 2003). In practice, some components would likely need more standardisation across settings than others, depending on evidence and theory. Furthermore, the process of local identification and implementation of effective strategies has been shown to be standardisable in its own right (for example, see Bond et al., 2001).

A functional approach towards implementation fidelity need not damage evidence-based practice but requires a different perception of what becomes evidence-based as the result of rigorous evaluation (a set of functions instead of a set of activities), and a substantially more rigorous effort of intervention theorisation to be a valuable alternative to current research practice. Fidelity monitoring at each local site would use common methodology to monitor the implementation of locally designed activities, and the extent to which core standardised functions are present. However, the literature is currently lacking any conceptual work on the distinction between the functions as opposed the specific form of social intervention programmes, and what a focus on functions when standardising programme components would look like in empirical studies. Such work is paramount in order to assess the extent to which this approach may provide a useful alternative towards conventional thinking about implementation fidelity, and to examine empirically if this can help the development and implementation of more effective interventions or not.

Developing a more detailed understanding of intervention adaptation

In the evaluation literature, the concept of adaptation of interventions has received much less attention than the concept of implementation fidelity, particularly in empirical studies (Moore et al., 2013; Ozer et al., 2010). Whereas fidelity has been conceptualised, operationalised, and measured for over two decades, the concept of adaptation has been treated as a black box for much of that period. Definitions of implementation fidelity as a lack of modification (adaptation) of manualised content testify to this (Elliott & Mihalic, 2004). However, adaptation has received increasingly more attention recently. An important development consists of researchers' increased recognition that not all adaptations damage fidelity, that fidelity and adaptations may occur simultaneously without damaging intervention outcomes, and may improve them in some cases (Durlak & DuPre, 2008; Dusenbury et al, 2005). This recognition of the potential value of adaptations for interventions' theories of change in addition to its mere practical value for the benefit of their dissemination has prompted the development of frameworks to guide practitioners in the process of adapting EBPs to new contexts (Lee et al., 2008). It has also led to more detailed analyses of adaptations. Specifically, authors have further developed our understanding of the ways by which interventions are adapted and the reasons practitioners and providers may have for making different types of adaptations. Hill and colleagues (2007) found that most providers involved in a community-based substance abuse prevention program reported that fidelity to curricula was important, but also reported adapting the program. These providers cited a shortage of time as the most frequent reason for prompting adaptation. This reason was particularly regarded as a main cause for *deleting* program content. However, participant group characteristics were the most mentioned cause for *changing* program content. Ozer et

al. (2010) conducted similar research, and found that teachers of substance abuse and violence prevention curricula were most likely to make changes to the instructional format of these programs (modification), integrate real-life experiences into the curricula (addition), or supplement them with additional resources (addition). The most important rationales for these adaptations were pedagogical and class management issues. Such research as described above has recently prompted an effort to develop a *taxonomy for understanding adaptations* (Moore et al., 2013). Moore and colleagues coded adaptations made by providers of ten EBPs in Pennsylvania based along three dimensions: fit (philosophical versus logistical), timing (proactive versus reactive), and valence (alignment with or deviation from a program's goal and theory). This three-dimensional conceptualisation of adaptation made it possible to assess the extent to which reasons for adaptation (caused by philosophical or logistical motives), and their timing (planned before implementation based on anticipated problems or occurring during the implementation process), affect the extent to which adaptations might be expected to fit the programmes' underlying theory of change. Moore and colleagues (2013) concluded that most adaptations (but not all) were made reactively as a result of logistical issues, and were not aligned with or deviated from programme theory.

1.4.6 Section Summary

Monitoring the implementation of programmes forms an important part of intervention research, as understanding programme implementation is essential to understanding interventions' effects, building up their evidence base, and to disseminate those that are successful. One of the key concerns in intervention- and implementation research is to establish and maintain implementation fidelity: to implement programmes as intended. To fulfil this objective, researchers typically

specify and standardise intervention components a priori, and subsequently monitor their implementation across sites. Sometimes, such activities are manualised and practitioners or other implementers are guided to adhere to programme manuals. A conflict arises when these or other stakeholders desire to adapt components of services to the needs of their clients, organisations, or contexts. Such adaptations are rule rather than exception and may threaten the fidelity with which pre-specified interventions are implemented. The academic debate revolving around this conflict is often referred to as the fidelity-adaptation debate, and devoted advocates can be found for the case of maintaining implementation fidelity as well as the case for leaving room for contextual adaptations. However, recent efforts show that the conflict between fidelity and adaptation might not need to exist. First, the standardisation of interventions might be conceptualised differently, avoiding the prescription of a fixed set of activities to practitioners or other implementers. If supported by a complete and detailed theory of change, it may be feasible to standardise interventions meaningfully in terms of their underlying theoretical change processes. When these change processes are specified explicitly, local implementers can map intervention activities that match their contexts to these change processes and implementation fidelity can be monitored as fit between these. This would facilitate the standardisation of interventions while leaving room for local (re-)invention or adaptation of intervention components. Second, researchers now recognise that the adaptation of interventions does not imply a loss of implementation fidelity per se. Adaptations are made for different reasons and during different stages of the adoption and implementation process. Whereas some adaptations may conflict with programmes' underlying theories (thus likely undermining their effects), others may be in line with or even reinforce these theories. At this point, too little empirical work has been conducted on

novel ways to conceptualise the standardisation of interventions and implementation fidelity in order to provide a feasible and sensible alternative to the standardisation of interventions' activities. Further research is thus needed to move the fidelity-adaptation debate forward.

1.5 Objectives and research strategy of this thesis

This thesis aims to make two main contributions to the improvement of social programmes and their evaluation. The first is to assist in the development and initial implementation of the Studio School model of secondary education in England by analyzing current implementation practices and providing recommendations for improvement of these in an effort to prepare the model for a future effectiveness evaluation. The second intended contribution is to advance the fields of implementation and intervention science by developing, testing, and discussing a new approach towards thinking about and monitoring of the concept of implementation fidelity. Specifically, the thesis aims to address the following objectives.

1. Specify a programme theory after its development but prior to its evaluation

The importance of explicit programme theories for evaluation has been discussed in this chapter, and will be further discussed in the next chapter. Programme theories are often not specified in detail or through a systematic process prior to evaluation, and this is particularly the case when programmes are not developed by evaluators. This thesis aims to make a contribution to the scientific literature on programme theory development by answering the following research question.

Is it possible to elucidate a clear theory of change of an intervention, where this has not previously been developed, based on the consensual accounts of non-academic developers?

This research question will be answered in Chapter II using a combination of qualitative and quantitative methodology, structured around a Delphi-inspired consensus process.

2. Conceptualise, operationalise, and empirically assess a new approach towards implementation fidelity

The tension between preserving implementation fidelity and adapting intervention programmes to fit the contexts in which they are implemented could be alleviated if the former focused less on their specific form. Standardising interventions by their functions has been suggested as a potential method for accomplishing this, but no rigorous efforts have been undertaken to clearly define and operationalise such an approach. Its practical utility for implementation and intervention research has therefore not been researched. Hence, this thesis seeks to explore these issues by providing an answer to the following research question.

Can the implementation of complex social interventions be measured validly and reliably in terms of the theoretical functions (next to the form) of their components?

In order to answer this research question, a conceptual framework for defining implementation fidelity from a function-focused perspective will be laid out in Chapter III, upon which a multi-stage measure development process is described in Chapter IV. In Chapter V, the research question is answered based on an assessment of the psychometric properties of the used instruments in the chapter, along with a discussion of the previous chapters.

3. Provide insight to Studio Schools regarding their current implementation practices

Studio Schools represent a new method of secondary education that has the potential to influence the lives of many teenagers. Although only an outcome evaluation can provide insights in the effects of the model, such an evaluation will lose much of its explanatory power without a thorough understanding of how schools work with some of its core components and to what extent these are being implemented with fidelity. To allow for the flexibility regarding tailoring of the model to each school's context that is advocated by the Studio School movement, implementation fidelity is monitored by function alongside form. To gain more insight in the dynamics between these types of implementation fidelity, the relationship between both concepts is also assessed. These objectives are summarised in the following research questions:

Comparing across schools, does the Studio School model currently appear to be implemented with fidelity defined by form and function?

This research question will be answered based on the primary implementation study conducted in Studio Schools that is reported in Chapter VI. Using quantitative implementation checklists (the measures developed in Chapter IV and V) completed by different groups of observers, an image is given regarding current Studio Schools implementation practices.

To what extent are form-focused and function-focused fidelity scores for project-based learning and personal coaching in Studio Schools correlated?

This question will be answered in Chapter VI by examining and discussing the correlation between the form- and function-focused fidelity scores collected during the primary implementation study.

1.6 Chapter summary

This chapter introduced the objectives of this doctoral thesis, and provided the background and context within which these were developed. It described England's complex education system that is rooted in a demand for diversity and designed by generations of policy makers with conflicting ideas. In spite of this, England and the UK as a whole face some serious difficulties with regards to the education of young people. Disengagement with education causes one child in five to fail to gain a qualification that equals at least Level 2, and attainment in English secondary schools ranks low amongst high-income European countries. In addition, a lack of priority devoted to nurturing students' non-cognitive skills leaves a large proportion of secondary school leavers in England insufficiently prepared for the world of work. Studio Schools are designed to tackle the twin challenges of a shortage of transferable employability skills of secondary school leavers and low levels of engagement with school experienced by a large group of young people, and are the social programme of focus of this thesis. The schools aim to offer a mainstream option of secondary education for students of all abilities who learn better when their learning is practical and contextualised in the real world. Although only an outcome evaluation can determine whether the model is effective towards achieving its goals, some important challenges inhibit such an evaluation to be valuable at this point. Specifically, as Studio Schools are founded and led locally, it is unclear whether different schools actually work with the same model. This is particularly pertinent as Studio Schools are actively encouraged to reinvent the model to suit their local circumstances and population. Therefore, the elucidation of the Studio Schools theory of change and an assessment of the fidelity with which this theory is implemented across schools should be the priority of an evaluation at this stage and the remainder of this thesis

focuses on these elements (objectives 1 and 3 of the thesis). The intended flexibility of the Studio School model makes its evaluation prone to face a tension between implementation fidelity and contextual adaptation of practices. However, important steps have been made in the evaluation literature to relieve this tension or even resolve it. Novel ways of conceptualising interventions and programmes may make room for more flexible definitions of implementation fidelity, and detailed research into adaptations reveals that not all adaptations have to undermine the original intentions of program designers. However, little empirical work has been done at this time to verify the assumptions that underlie such efforts and increase our understanding of how implementation fidelity and contextual adaptation may coexist. This is particularly relevant for innovations such as Studio Schools, that rely on locally founded and led schools but that aim to provide a distinguishable brand of schools that work with the same core model. Hence, this thesis will focus in particular on improving the methodology for allowing flexibility in evaluations of complex social interventions (objective 2). First, a detailed programme theory of the Studio School model needs to be developed. This process as well as its output is reported in the next chapter.

Chapter II: Specifying the programme theory of Studio Schools

2.1 Introduction

Section 1.5 of the previous chapter stated the objectives of this DPhil thesis. In the current chapter, the first of these objectives is addressed: the elucidation of a programme theory after a social programme is developed but before it is evaluated. Specifically, it reports the process of developing the programme theory for Studio Schools and summarizing it in a logic model. This chapter resulted in the following publication:

Van Urk, F.C., Grant, S., & Bonell, C. (2015). Involving Stakeholders in Programme Theory Specification: Discussion of a Systematic, Consensus-Based Approach. *Evidence & Policy*. DOI: <http://dx.doi.org/10.1332/174426415X14474260456850>.

2.1.1 Research question

The objective of this study was to develop a logic model for Studio Schools that would guide the remainder of this doctoral project as well as any future research on Studio Schools. Additionally, the first research question of the doctoral project is addressed.

Is it possible to elucidate a clear theory of change of an intervention where this has not previously been developed based on the consensual accounts of non-academic developers?

2.2 Programme theories and stakeholder involvement

2.2.1 The importance of explicit programme theories

The importance of theory-based intervention design and evaluation has been recognised for decades (Chen & Rossi, 1983; Michie et al., 2005; Rogers, 2007;

Weiss, 1997). Indeed, the use of explicit programme theory –a description of the causal chain of activities intended to produce a positive intervention outcome (Fraser et al., 2009)– has several important benefits. First, there is increasing evidence that theory-based interventions may be more effective than those interventions not explicitly based on theory (Kinmonth et al., 2008). This may largely be explained by the development of theory-based interventions on a grounded understanding of behaviours and the contexts in which these occur and thus a better informed approach towards targeting these (Glanz & Bishop, 2010; Noar et al., 2008). Furthermore, the use of explicit theory may expose weak links or potential conflicts in the hypothesised change process that should be investigated empirically and potentially revised (Moore et al., 2014). Additionally, it can aid in the systematic assessment and prevention of potential adverse effects of programmes (Bonell et al., 2015). Recognising these benefits, recent guidance on developing complex interventions suggests that explicit theories of change be used to inform the design of interventions. (Campbell et al, 2007; Craig et al, 2008; Glanz & Bishop, 2010, Moore et al., 2014). Second, explicit programme theories improve the evaluations of programmes. Pawson and Tilley (1997) argue that all interventions are by definition ‘theories incarnate’ because they initiate a particular causal process leading to particular impacts. As a result, the use of explicit programme theory greatly improves the informative power of evaluation research. This is particularly pertinent for complex social interventions. Prior to a full-scale evaluation, clarifying how prescribed intervention components are supposed to facilitate mediating change processes and how the process is moderated by contextual factors can provide important information about the intervention for evaluation design (Craig et al., 2008). Researchers furthermore stress the importance of using explicit theory within evaluation because it facilitates the contribution of trials to the testing

and refinement of generalisable programme theories across contexts and populations, instead of merely specific intervention programmes (Bonell et al., 2012; Michie & Abraham, 2004; Michie & Prestwich, 2010). For example, the school-based INCLUSIVE intervention to decrease bullying consisted of a set of explicitly modelled standard structures and processes that aimed to facilitate changes in school ethos and practice. Whereas these components would be implemented the same across schools, the activities delivered to students as a result of organisational changes were allowed to vary between schools (Bonell et al., in press). Further benefits of using explicit theories for evaluation include an improved understanding of why interventions are effective or ineffective, and potentially a greater influence of evaluation results on popular opinion and policy (Weiss, 1995). Finally, a clear programme theory facilitates the successful implementation of interventions. Through explicit modelling, the essential resources for a programme's success may be clarified up front so that these can be put in place prior to implementation (Cooksy et al., 2001; Huhman et al., 2004). It may also uncover and resolve different understandings of intervention models according to different stakeholders (Moore et al., 2014). Researchers also recognise the role of explicit theory in specifying the core components of programmes that should be implemented with fidelity as a requirement for effectiveness (Hawe et al., 2004a; Gardner et al., 2010).

In spite of recommendations, studies often fail to explicitly apply and test theory, and in some cases researchers or evaluators fail even to refer to programme theory (Gardner et al., 2010; Noar & Zimmerman, 2005; Trifiletti et al., 2005). For example, a review of reporting practices of smoking cessation interventions found that less than half of the components that formed part of programme manuals were described in study reports published in journals (Lorencatto et al., 2012). This results in readers

gaining limited insight into the intervention processes that were tested in studies. This does not imply that such studies evaluate interventions or programmes that have no basis in theory: even when the theory of change and the structure of a programme are not specified explicitly, they are often present implicitly (Chen, 1990). Nevertheless, to systematically guide a process- or outcome evaluation, as well as programme implementation, an explicit theory of change is preferable (Connell & Kubisch, 1999; Eldridge et al., 2005). Given the intricacies of programme theory for some interventions, the specification of logic models has been recommended for complex interventions. Logic models are diagrammatic representations of the relationships between an intervention's resources, activities and intended outcomes and often make them easier to comprehend by a variety of stakeholders (Cooksy et al., 2001; Kellogg Foundation, 2004; Trevisan, 2007). This can be particularly helpful when engaging in evaluative efforts with programme developers that are not used to articulating explicitly why the dissected elements of a programme achieve its goals (Monroe et al., 2005). Although informed by implicit theories present in the minds of their developers, interventions designed by non-evaluators often lack an explicit theory that specifies how they work and how hypothesised change process are supported by evidence (Evans et al., 2014). Because such interventions form the majority of actually implemented social programmes, explicating their theories can make an important contribution in bridging the gap between research and policy. In order to evaluate such programmes rigorously and usefully, it is imperative that their underpinning implicit theories are rendered explicitly prior to evaluation.

2.2.2 Elucidating implicit programme theories and building logic models

It is not uncommon for evaluators to specify the programme theory that underlies an intervention after its design but prior to evaluation (Connell & Kubisch, 1999). When

developing programme theory and building logic models, Patton (2008) distinguishes between three overarching approaches: deductive, inductive, and user-focused. *Deductive approaches* involve researchers using previous empirical research and dominant theories from different fields to inform programme theory development. This is typically how researchers develop programme theories before a programme is implemented. If executed rigorously, these approaches hold the advantage of being rooted in systematic scientific enquiry. However, they do not allow for the integration of implementation experiences as well as contextual factors into their models. When using an *inductive approach*, researchers generate programme theory by observing the programme as it is being implemented, and reviewing documentation (Oosthuizen & Louw, 2013; Savaya & Waysman, 2005). Although not always grounded on a large body of prior research and theoretical frameworks, inductive approaches allow researchers to base their programme theories on data from the world of actual practice. For example, Savaya and Waysman (2005) reconstructed the programme theory of a contact-resuming training programme for parents who had been separated from their children. Specifically, they assembled logic model components based on government documents describing programme expectations, constructed a coherent set of items, and finally examined how these could be linked based on the documentation. Within *user-focused approaches*, evaluators collect information from programme staff through interviews or surveys, and use the data to construct a programme theory. When employing such approaches, some evaluators actively involve a programme's practitioners, stakeholders, or purveyor organisations in elucidating an implicit theory of change through interviews and consultation processes (Christie & Alkin, 2003; Julian, 1997; Oosthuizen & Louw, 2013; Sullivan et al., 2002). When Yampolskaya and colleagues (2004) elucidated the programme

theory of a child mental health programme, for example, programme staff members and coordinators were involved in group brainstorming and a rating process to arrive at a logic model that was subsequently reviewed again by participants. Like inductive approaches, such participatory efforts are not always grounded in rigorous scientific inquiry. However, they can be useful in elucidating programme theories of interventions that have previously been left implicit by their developers. In order to obtain an accurate understanding of how developers and implementers perceive their programmes, inductive user-focused approaches are required because a deductive approach –however valid– might result in a different intervention. In addition, these may increase stakeholder commitment to these theories and subsequent evaluations. Unfortunately, the specific methodological processes for elucidating programme theory are rarely explicitly reported. Consequently, it is often unclear exactly how information gained from deductive, inductive, and user-focused approaches result in a final product (e.g., a logic model). This lack of transparency makes it difficult for practitioners, evaluators, and other stakeholders to assess the validity of programme theories, and prevents the accumulation of knowledge on methods to elucidate these theories for future research and evaluation.

2.2.3 The Delphi method

Various methods could be used within user-focused approaches towards elucidating programme theories. Methods that facilitate the development of consensus may be useful to offer guidance to stakeholders with regards to their agreement or disagreement related to specific parts of their programme. The Delphi method is a consensus development technique that involves facilitated, structured interaction amongst individual stakeholders informed on a given topic area. A series of questionnaires are presented in an iterative, multi-round process to systematically

collect, aggregate, and present to a panel of stakeholders the group's opinions or judgments on specific questions and issues related to the topic area of interest (Brown, 1968; Hasson et al., 2000; McKenna, 1994). Three strengths of the Delphi technique are:

- The absence of face-to-face contact, which allows interaction of a geographically sparse group of stakeholders while avoiding potential pitfalls of in-person discussion (e.g., dominance of discussion by overly assertive individuals, reticence by introverted individuals);
- The formal feedback of group views to participants after each questionnaire, which facilitates reconsideration of previously expressed views after comparison to others' responses;
- The explicit aggregation of individual responses, which provides an auditable trail of how an end product is the result of stakeholder consultation (Murphy et al., 1998).

Its pragmatic utility has made the Delphi a popular method in the social and health sciences. The Delphi method has been used for a variety of purposes, including the determination of core outcomes to measure in randomised controlled trials (RCTs), setting research priorities in educational technology, and developing reporting guidelines for health research (Hopewell et al., 2008; Pollard & Pollard, 2004; Sinha et al., 2011). In the current study, we elucidated the programme theory of Studio Schools. In this study, inductive and user-focused approaches were combined in order to maximise the engagement of core stakeholders and to utilise their experience in elucidating a programme theory that would likely reflect real world implementation of the model rather than an ideal representation. Specifically, the Delphi method was included into an explicit and transparent methodological process of incorporating stakeholder input in programme theory development. Because it was expected that participants would be unfamiliar with programme theory articulation, the logic model was used as a framework to provide a comprehensible structure to the process.

2.3 Methods

A two-round Delphi process was integrated into a systematic effort to involve stakeholders in elucidating the theory of change of Studio Schools. This effort took place from December 2012 to August 2013. Participants were emailed a link to the electronic questionnaires at the beginning of each round (see Appendix A). This study was reviewed and subsequently approved by the Departmental Research Ethics Committee of the Department of Social Policy and Intervention, University of Oxford (Ref: 2012-13_06).

2.3.1 Participants

To be considered as stakeholders eligible for the current study, participants had to be i) originators of the Studio Schools approach, ii) responsible for its promotion and facilitating its implementation, or iii) implementers of the model in schools (i.e., principals of Studio Schools). Ten individuals (four female, six male) were approached to take part in the study, based on their nomination during preliminary interviews with both the originators of the Studio Schools approach as well as members of the Studio Schools Trust. Three of these had shared some responsibility for developing the original Studio Schools concept. A further three participants were current members of the Studio Schools Trust and responsible for the model's dissemination across England. Finally, four participants were current principals of Studio Schools. All approached individuals agreed to participate and provided written consent.

2.3.2 Preliminary interviews

Prior to commencing the Delphi process, semi-structured interviews were held with the three participating originators and developers of the Studio Schools approach and

two members of the Studio Schools Trust. These interviews provided an understanding of the background and context of the Studio Schools development process, and helped identify potential Delphi participants. The interview data were also used to inform the development and analysis of the Delphi questionnaires.

2.3.3 Delphi process

Round One

The Delphi process started with an ‘open’ first round. That is, participants were sent an open-ended questionnaire asking them to nominate, in their own words, the most important features of the Studio School model (see Appendices A1 and A2). With no explicit logic model to draw on, an open first round was deemed important so that the theory of change could be elucidated by stakeholders of the Studio School model rather than being developed and later imposed by the researcher. Participants had four weeks to complete Round One. At the start of the questionnaire, participants were provided an introduction explaining the concept of a logic model and the purpose of the study. The rest of the questionnaire was structured by different elements that form part of a logic model (‘objectives’, ‘inputs/resources’, ‘activities’, ‘outputs’, ‘proximal outcomes’ and ‘distal outcomes’ of the programme). Participants were asked to list all components of the Studio School model, providing a short explanation (100 words) of what is meant by each feature and about the importance of each feature. As it was assumed that not all respondents would be familiar with the elements of a logic model, one hypothetical example of a Studio School feature was given per logic model element (e.g., project-based learning for ‘activities’). There was a specified limit (n=8) of items respondents could suggest per logic model element, informed by initial interviews that suggested this was an appropriate number. Two researchers independently analysed and integrated responses into a list of non-

redundant items, structured by the elements of a logic model (see Supplement C). Disagreements between the researchers were resolved following discussion, after which a final list of items was constructed. This list was sent to all participants prior to their rating of items in Round Two. It was accompanied by each participant's original responses and a description of how each was integrated into the list, so that participants could understand the process leading to the development of the Round Two questionnaire and add or rephrase items if deemed necessary. Participants were not provided with the individual responses of any other panel member: they only saw others' responses in aggregate form.

Round Two

The second round consisted of respondents filling out a closed-response questionnaire and took nine weeks for all participants to complete (see Appendix A3). This questionnaire contained the consolidated list of items that were identified in Round One, and it was structured by the same logic model elements as the Round One questionnaire. Respondents were asked to rate the relevance of each nominated item to the Studio School model on a 1-10 Likert-scale, with a rating of '1' representing a judgment of an item as 'not at all relevant', and a rating of '10' representing a judgment of an item as 'highly relevant'. In addition, participants were given the opportunity to provide written clarification for their rating of each item. Ratings were summarised as simple descriptive statistics (medians, modes, lowest and highest values). Ratings were then analysed using adapted criteria from the RAND/UCLA appropriateness method manual, which classifies items in accordance with i) their median panel rating and ii) some measure of the dispersion of panel ratings, which is taken as an indicator of the level of agreement with which the ratings were made (Fitch et al., 2001). Upon receipt of participants' responses, the decision was made to

use the ‘A7S’ procedure for classification of items in order to preserve a strict definition of consensus while minimising the influence of extreme ratings by outliers. According to this definition of agreement (Fitch et al., 2001), an item was considered to have consensus if, after discarding the most extreme high and the most extreme low rating, the remaining ratings all fell within the same ‘range category’ of 1-4, 5-7, or 8-10. Adapting this procedure for our panel size of 10, initially items could be classified as included, excluded, or uncertain (see Table 2.1 for a description).

Table 2.1: A7S Criteria for the classification of items after Delphi Round Two ratings by participants (adapted from Fitch et al., 2001)

Item rating after discarding highest and lowest rating	Decision
Value of 8 or higher by all panel members	Included
*Median value ranges from 5-7 *Values do not all fall into a single range category (1-4, 5-7, 8-10)	Uncertain
Value of 4 or lower by all panel members	Excluded

Using Delphi results to create a logic model

Items classified as ‘uncertain’ from the quantitative decision rules above were then either included or excluded from the logic model by the doctoral candidate, based on data obtained from: i) numerical ratings and qualitative responses provided by participants in the first- and second rounds of the Delphi process, ii) semi-structured interviews that were conducted with some panel members prior to the first round, and iii) documentation provided by the Studio Schools Trust and the Studio Schools online network. Items were included if explicit support for their classification as a necessary feature of the Studio School model could be found in at least one of the sources above. Items were excluded if such support was absent from all of the sources above. Following this selection process, the first doctoral candidate created a logic model linking the included items. Links (relationships) between items were again

based on explicit support for such relationships in the same sources used for the inclusion or exclusion of ‘uncertain’ items described above: i.e., Delphi participant responses, semi-structured interviews, and official Studio School documents. The first author then wrote a report describing the Delphi process and its rationale, and depicting the proposed Studio Schools logic model. All participants and the second coder (another doctoral candidate) were sent the report, given forms to provide feedback on the Delphi process and the proposed logic model, and asked to suggest changes if they felt these were needed. Feedback led to a revised draft logic model. The complete post second round process of including and excluding items, as well as the drafting and circulation of reports to participants and the second coder, took the doctoral candidate approximately 6 working days to complete.

2.4 Results

Nine participants completed the Delphi Round 1 questionnaire and nominated items for inclusion in the Studio Schools logic model. Ten participants completed the Round 2 questionnaire, rating and commenting on proposed logic model items (see Figure 2.1).

2.4.1 Delphi process

Participants nominated a total of 314 unique items in Round One. The majority of qualitative comments explained why and how nominated items fit into the overall Studio School model (e.g., how an ‘objective’, ‘input’, or ‘activity’ is important to the development of students or facilitating other components of the model). In some cases, participants clarified or elaborated on their meaning of a nominated item; for example, what they specifically meant by an item in the context of Studio Schools (for a complete list of all nominated items including explanations and elaborations by

participants, see Appendix A2). After independent integration of items into a non-redundant list by two researchers, a total of 59 items were rated by participants in Round 2. Upon receiving the integrated item list and their original responses prior to the start of Round 2, participants did not express any concerns regarding the validity of the list. Participants rated 59 items in Round Two. Median ratings for each item ranged from 6 to 10, though individual participant ratings spanned the full range of 1 to 10. When commenting on ratings, participants mostly elaborated on relatively low ratings. In some cases, participants explained why an item was not essential to the Studio School model, or was only relevant to some schools.

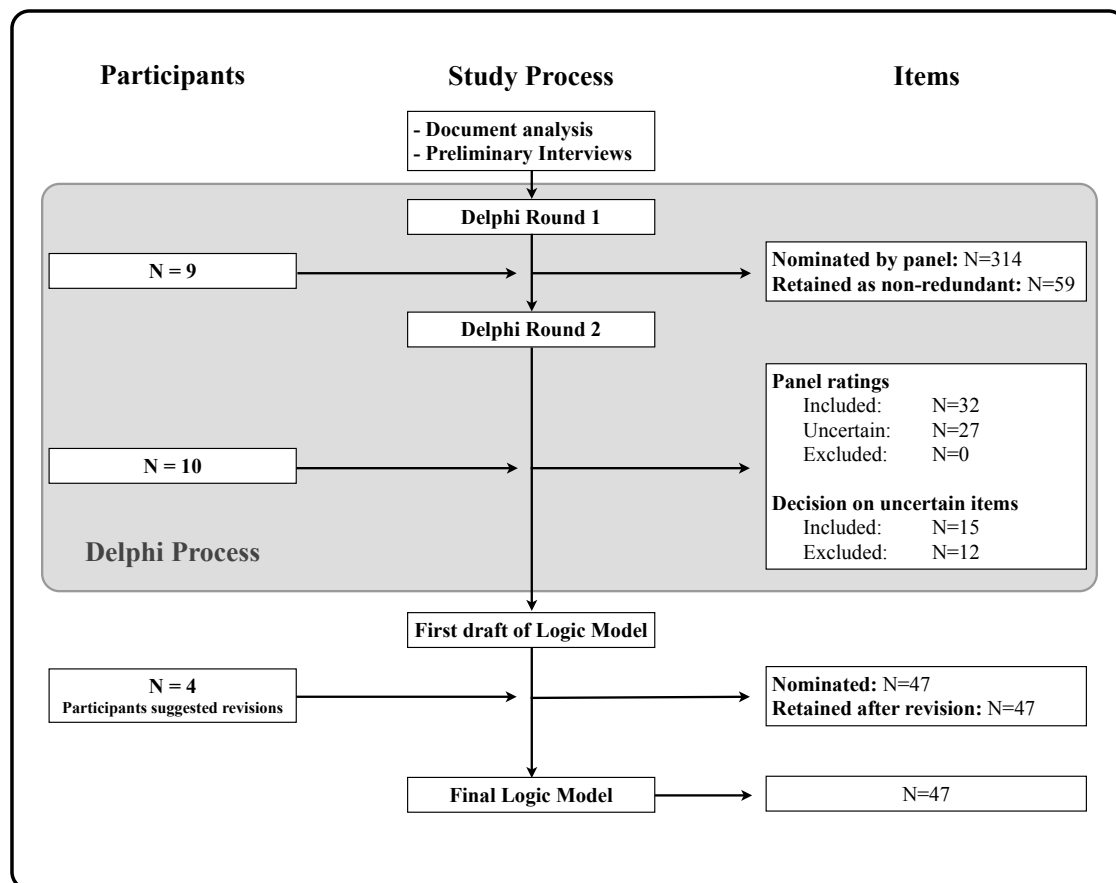


Figure 2.1: Flow of participants and items through the study

Some participants were also concerned that, although relevant, an item might not be unique to the Studio School model. Finally, participants occasionally explained that their relatively low rating for an item was the result of the item being phrased either

slightly off-point or too exclusively for some students at Studio Schools (for a complete description of the Round Two rating process, see Appendix A4). Table 2.2 presents the classification of items after Round Two. Overall, 32 items were classified as included based on participants' ratings, according to the adapted A7S procedure (Fitch et al., 2001). A further 27 items were classified as uncertain. No items were classified as excluded at the end of the second round.

Table 2.2: Classification of items based on ratings in Delphi Round Two

Item category	Rated as included (all ratings 8-10)	Rated as uncertain (4 < median < 7 OR ratings do not fall in same category)	Rated as excluded (all ratings 1-4)
<i>Objectives</i>	3/7	4/7	0/7
<i>Inputs</i>	10/13	3/13	0/13
<i>Activities</i>	5/15	10/15	0/15
<i>Outputs</i>	2/6	4/6	0/6
<i>Proximal Outcomes</i>	8/10	2/10	0/10
<i>Distal Outcomes</i>	4/8	4/8	0/8
Total	32/59	27/59	0/59

2.4.2 Creation of the Studio Schools logic model

Of the 27 items classified as uncertain at the end of Round Two, 15 items were included by the doctoral candidate based on support found in Delphi participant responses, semi-structured interviews, and official Studio Schools documentation. The remaining 12 items were excluded because no explicit support for these was found in these sources. An example of the process of final classification of 'uncertain' items nominated under 'outputs' is presented in Table 2.3 (for a complete overview of the decision process on items classified as uncertain after Delphi Round Two, see Appendix A5). Once a list of included items was defined, and the researcher developed a first version of a logic model, some items were integrated under common themes to preserve a logically coherent model. For example, some core activities such as 'project-based learning' and 'personal coaching' of students were integrated under

the theme ‘school-based curriculum’. After the report describing the Delphi process and the first draft of the logic model was distributed to all 10 participants, 4 of these requested some items to be disintegrated from common themes, as these were considered distinct operating functions of the model. These participants represented key stakeholders for the subsequent evaluation phase. Revisions were made accordingly and then the logic model was re-circulated. Participants expressed no concerns regarding the final version of the logic model (see overview on page X).

Table 2.3: Example of final decision process on inclusion and exclusion of items.
Example items represent the six outputs rated by the panel in Delphi Round Two

Item	Description	Median	Range	Classific.	Decision	Reason
1	Work experience	9	8-10	Included	Included	Ratings
2	Positive social networks	9	8-10	Included	Included	Ratings
3	Personalised curriculum	10	7-10	Uncertain	Included	<i>‘One of 7 essential elements of a Studio School’ (SST, 2012)</i>
4	Appropriate academic qualifications	10	7-10	Uncertain	Included	Low panel ratings were due to concerns related to item description, which was revised
5	Portfolio	8.5	7-10	Uncertain	Excluded	No support found in consulted sources
6	Extra-curricular achievements	7.5	5-9	Uncertain	Excluded	No support found in consulted sources

2.5 Discussion

2.5.1 Summary of results

This study aimed to explore and demonstrate the possibility of a transparent, systematic, and user-focused approach for developing an explicit programme theory for a complex social programme based on the input of its stakeholders. Specifically, the present study provides an example of elucidating a logic model for a novel educational approach—Studio Schools—via active and systematic involvement of

those involved in creating, promoting, and implementing it. An iterative consultation method was incorporated in the form of a modified Delphi process, integrating participants' quantitative ratings and qualitative responses on components of the Studio School model into a coherent conceptual framework. This initial logic model is underpinning a subsequent formative evaluation of the Studio School approach; as such, it is not intended to be a definitive logic model to which each individual school must strictly adhere. Nevertheless, it specifies the core theoretical assumptions of the current Studio School model, which can be used to inform theory-driven evaluative research.

2.5.2 Strengths of this study compared to other approaches

When researchers seek to evaluate interventions of which they were not the developers, inductive approaches are needed to ensure that specified models represent the true intentions of developers and implementers. The use of a participatory, systematic, auditable, and replicable method that is founded on explicit consultation with stakeholders may be of use to others seeking to engage in similar efforts. Including a user-focused approach in the form of a modified Delphi process as part of the development of programme theory in combination with other commonly used inductive approaches—such as face-to-face interviews and document analysis—has several advantages over more deductive approaches. First, although the item inclusion process was to some extent influenced by the researchers, items nominated by participants in an open first round and subsequent rating process formed the basis of the model. Moreover, participants received explicit descriptions of all parts of the process and were given the opportunity to provide feedback at all stages. The resulting logic model therefore likely describes Studio Schools' programme theory as viewed by its main stakeholders.

The Studio Schools Logic Model

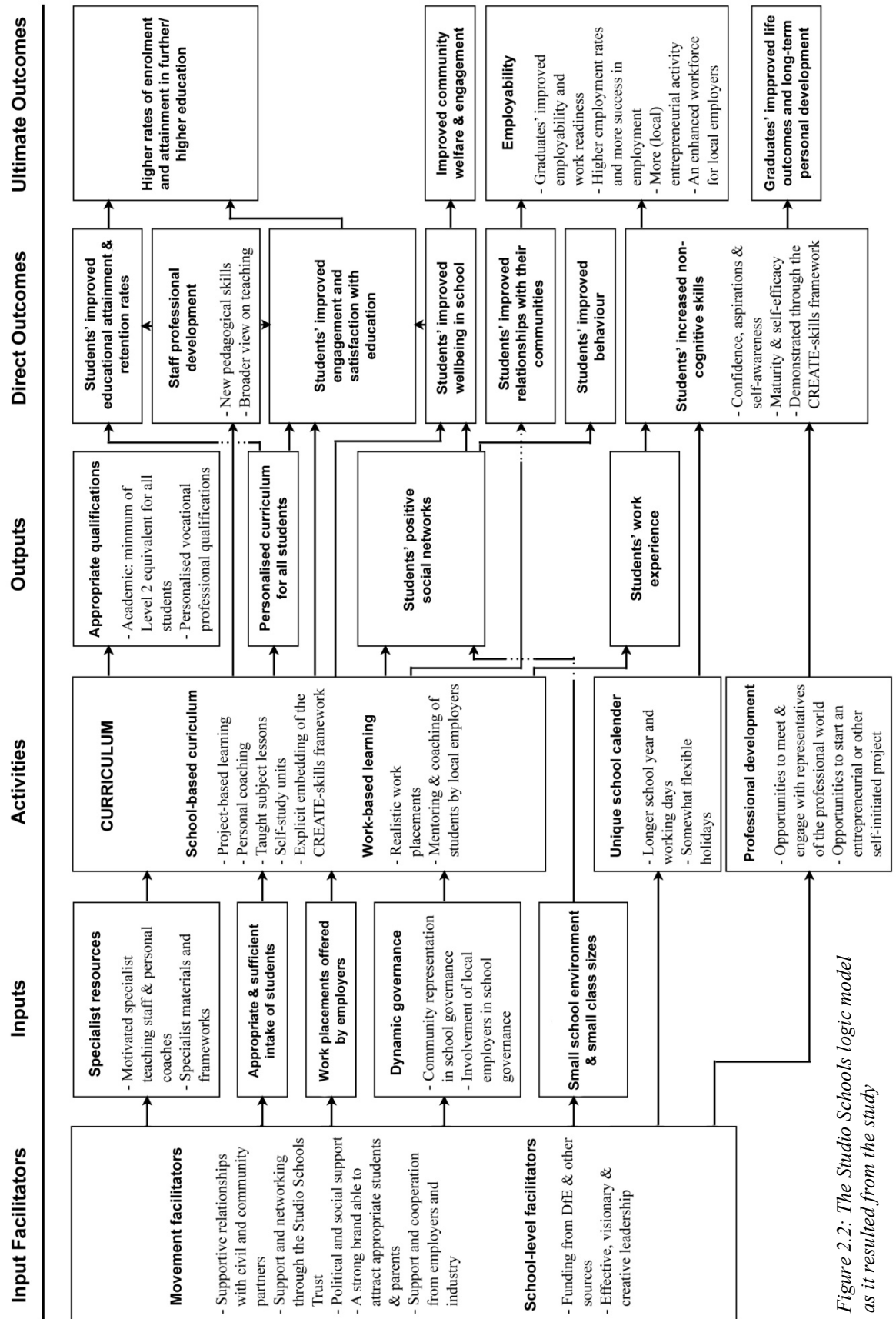


Figure 2.2: The Studio Schools logic model as it resulted from the study

This decreases the risk of incorrect interpretation by the researchers and underpins the intention of increasing buy-in by stakeholders in the conceptual framework underpinning future evaluations. Had the end product relied solely of the researcher's judgment based on analysis of programme documents or common social science theories, as is usually done within inductive and deductive approaches, this would have likely not been the case. Even user-focused approaches that involve stakeholders extensively in the programme theory development process often involve them after a preliminary model, based on researcher judgement, has been drafted and circulated (e.g. Christie & Alkin, 2003; Savaya & Waysman, 2005; Renger, 2011). Second, the absence of face-to-face interaction during an online Delphi process, in combination with the structured and anonymous feedback of participants' responses, aims to decrease the risk that dominant stakeholders impose their views on other participants. This increases the likelihood that the resulting logic model more accurately represents the views of all involved stakeholders. Addressing this risk is particularly important when some stakeholders hold positions of professional power over other participants, and when the results of subsequent evaluations could impact operations of intervention implementation. Previously, user-focused approaches to programme theory development such as structured concept mapping (Yampolskaya et al., 2004) and the Nominal Group Technique (Julian, 1997) all included in-person interaction between stakeholders, leaving it unclear how this may have influenced their results.

2.5.3 Limitations

Limitations of the taken approach

The decision to integrate a Delphi process into a larger, systematic user-focused approach to logic model development also carries some limitations. First, similar to previous user-focused efforts, the strong reliance on stakeholders' input in this study

has resulted in a model that was heavily driven by non-evaluators. The limited experience of participants with evaluative science may have resulted in the inclusion of items that were not essential to actual programme success or in the exclusion of items that were. A deductive approach that relied more heavily on rigorous scientific inquiry –for example through a systematic assessment of relevant literature– could have possibly led to a more compact model that would contain the mere core components of Studio Schools. This limitation was taken into consideration by the researcher while drafting the logic model, but there is a chance that evaluation resources could be allocated sub-optimally as a result of focusing on the wrong programme components. Second, although the absence of interaction between participants resulted in the previously discussed strengths, it carries the limitation that it prevents the synergy between participants’ accounts of a model that in-person interaction can offer as well as a good understanding of discrepancies in opinion and the reasons therefore (Murphy et al., 1998). Additionally, in-person interaction between the researchers and participants could have actually increased participants’ understanding of evaluative science. However, with consideration to the composition of the Delphi panel of and resulting potential power dynamics, it is argued here that the benefits created by the absence of face-to-face interaction outweigh these limitations.

Additional limitations specific to this study

The study is also subject to some additional limitations. Primarily, only a select group of stakeholders (N=10) were invited to participate in the Delphi process. Participants were purposefully sampled based on nominations by the SST—the main promoter of the Studio School model, which indicated that no further individuals were eligible for nomination. Participants included stakeholders who were responsible for the original

concept, expert staff of the SST, and a sample of school principals. It is unclear how a larger sample size would have affected the logic model and its components, and whether those nominated to participate are more likely to share views compared to other potentially eligible members of the stakeholder population. For example, it is unclear if the constructed model would have looked differently had Studio School teachers and students been included in the panel in addition to those engaged in policy and school management. It is also worth noting that, in analysing Round Two results, the adapted A7S procedure of excluding extreme ratings resulted in a larger number of items classified as 'included' by the end of Delphi Round Two than would have been the case without excluding extreme ratings. Thus, there are potentially more items in the logic model than would have been the case had extreme ratings not been discarded. Another potential limitation is posed by the decision not to extend the Delphi process to a third round. Instead of asking stakeholders to rate items that were classified as 'uncertain' for a second time and to have them link logic model components, the researcher was responsible for the decision to include or exclude uncertain items and then link items in the logic model. To compensate, participants were asked for their feedback on the integrated logic model that resulted from this decision-making process. The possibility exists that participants' 'passive' approval of the constructed model yielded a different set of included items than re-rating of items in a third round. Researchers or evaluators considering the use of the Delphi method to elucidate programme theory should carefully consider the time (in this case an approximate total of 18 combined working days by the research team) and resources needed to follow-up with participants for multiple questionnaire rounds. Finally, it should be stressed here that the approach described here aims to render consensual and explicit what might be contested and implicit about a programme

theory of change. It does not aim to enhance this theory. It might be that in some cases, theories of change are logically deficient or are contradicted by existing social science theories or empirical evidence. As realist evaluators have highlighted, programme theories may also inadequately theorise how intervention mechanisms interact with context to determine intervention outcomes (Bonell et al., 2012). In such cases, the process reported in this chapter might need to be followed by further work with programme developers to enhance programme theory and modify the programme itself.

2.6 Conclusion

Is it possible to elucidate a clear theory of change of an intervention where this has not previously been developed based on the consensual accounts of non-academic developers?

This study confirms the feasibility of specifying and reaching agreement on a social programme's underlying theoretical assumptions after intervention design but prior to evaluation. Making such assumptions explicit is key to conducting useful evaluations—such as process evaluations within effectiveness trials—and this explication of programme theory should be done regardless of stage of implementation and evaluation. Second, this work demonstrates that a systematic, inclusive methodological process can be employed to elucidate implicit programme theories. This is likely to increase the credibility of results and buy-in from stakeholders involved in the process, as well as other key partners in subsequent programme evaluations. Future studies undertaken to specify programme theories prior to evaluation are likely to benefit from use of methods like the Delphi process

that promote systematic techniques and transparent reporting. These may include other explicit methods for building consensus (e.g., the nominal group technique, which includes more face to face interaction), or a series of well-reported interviews with developers that investigate how specific social and behavioural theories informed their programme.

2.7 Chapter summary

This chapter discussed the process through which the theory of change that underlines the Studio School model was developed. Its main product is a logic model for Studio Schools that can guide implementation monitoring during the remainder of the doctoral project. Because a substantial number of schools were already open and working with the model without a clearly specified theory of change, the latter had to be specified some time after the programme had been developed. An additional objective was therefore to assess if integrating the Delphi method in this process could facilitate stakeholder participation in post-hoc logic model development and support for the use of the logic model in future evaluations. A Delphi process was incorporated into a larger effort to engage stakeholders in the elucidation of a logic model for Studio Schools. Participants (N=10) included representatives of the organisations that founded the model (or lead its dissemination), and Studio School principals. This process finally led to the creation of a logic model that was circulated to participants for comment and approval, and that was used as a framework to inform the implementation measures and evaluation discussed in the following chapters. Additionally, this study demonstrates the feasibility of a new, replicable, and auditable method for researchers to engage stakeholders in the elucidation of an existing education model after its design but prior to its evaluation.

Chapter III: Approach towards monitoring implementation fidelity in Studio Schools

3.1 Introduction

As became clear from Chapter I, understanding the implementation of social programmes is essential for understanding their effects. In Chapter II, the first step to monitoring the implementation of Studio Schools was undertaken, since an explicit theory of change is a prerequisite for successful implementation monitoring. The next step towards determining the fidelity with which the Studio School model is implemented within schools is to select appropriate measures based on its theory of change.

3.1.1. Difficulties with defining implementation measures for Studio Schools

Studio Schools do not work with manuals and no previously defined set of implementation measures is available for monitoring their implementation. Therefore, implementation measures have to be defined specifically for this project. In line with common practice in implementation monitoring, the same set of measures – representing programme components– should be used consistently across all participating schools in order to compare like with like (Gresham et al., 1993; Dusenbury et al., 2005). However, individual Studio Schools need to respond to different challenges imposed by their local contexts, and will often even need to tailor pedagogic strategies to individual students in order to secure their personalised approach to learning. Hence, the Studio Schools Trust actively encourages Studio Schools to tailor the model to their local needs and context (see paragraph 1.3.3 of Chapter I). This approach indeed indicates that Studio Schools are a complex –versus simple or merely complicated– intervention that responds to complex social issues

(Glouberman and Zimmerman, 2002). Table 3.1 presents a useful overview of this classification, and shows that the complex nature of Studio Schools and the issues they respond to present a potential threat to the validity and utility of comparing implementation across schools with a set of standardised measures. Therefore, Studio Schools offer a perfect example of the tension between implementation fidelity and contextual adaptation for complex interventions that is discussed in paragraph 1.3.4 of the first chapter, and a revisit of what constitutes implementation fidelity in this context is necessary in order to determine meaningful implementation measures for Studio Schools.

Table 3.1: Distinction between simple, complicated, and complex interventions
(Glouberman & Zimmerman, 2002)

Simple	Complicated	Complex
<i>Following a recipe</i>	<i>Sending a rocket to the moon</i>	<i>Raising a child</i>
The recipe is essential	Formulae are critical and necessary	Formulae have a limited application
Recipes are tested to assure easy replication	Sending one rocket increases assurance that the next will be OK	Raising one child provides experience but no assurance of success with the next
No particular expertise is required, but cooking expertise increases success rate	High levels of expertise in a variety of fields are necessary for success	Expertise can contribute but is neither necessary nor sufficient to assure success
Recipes produce standardized outputs	Rockets are similar in critical ways	Every child is unique and must be understood as an individual
The best recipes give good results every time	There is a high degree of certainty of outcome	Uncertainty of outcome remains
Optimistic approach to problem solving	Optimistic approach to problem possible	Optimistic approach to problem possible

3.2 Revisiting implementation fidelity

Paragraph 1.4.4 of Chapter I demonstrated that, although practiced frequently and often sensibly, there are risks involved in standardising the components of complex interventions, affecting their dissemination as well as their effects. For an evidence-

based intervention to be replicable across settings, however, some standardisation of what constitutes the intervention is necessary and the above does not imply that no standardisation of any kind is possible. Researchers have argued that it is not the need for standardisation that is a threat to their dissemination, but *how* interventions are commonly standardised (Hawe et al., 2004a). Good evidence-based practice successfully integrates clinical or practitioners' [context-informed] judgement with the best external evidence (Sackett et al., 1996; Gibbs & Gambrill, 2002). In order for the evidence-based paradigm to be successful, attempts to generate and disseminate high quality evidence must consider the need for this integration. According to some researchers, this is insufficiently done, which may lie at the root of the fidelity-adaptation debate. Carol Weiss' distinction between 'programmatic theory' and 'implementation theory' articulates this. Programmatic theory deals with the *mechanisms* that intervene between the delivery of program service and the occurrence of outcomes of interest (Weiss, 1997). In order to test a programmatic theory –or theory of change– researchers' primary concern should be to test whether the theorised mechanisms of change occur when the implemented activities are consistent with those specified in the theory. Implementation theory focuses on how the program is carried out and assumes that if a program is conducted as planned, with sufficient quality, intensity, and fidelity to plan, the desired results will be forthcoming. The conceptualisation of interventions is often heavily based on implementation theory, rather than programmatic theory. *Standardisation of form* is the result of this, whereas *standardisation of function* (the necessary role of activities in the intervention's theory of change) might be a more appropriate method in order to resolve the tension between standardisation and adaptation. This idea is not new. The problem of suboptimal specification of theoretical change mechanisms as part of

intervention evaluations has been pointed out numerous times before (Chen & Rossi, 1983; Connell & Kubisch, 1998; Pawson & Tilley, 1997; Rogers, 2007; Lee et al., 2008). By some, the assumed fact that ‘we typically do not know which components of a program may be responsible for [effects]’ is cited as a key argument for the standardisation of interventions’ activities in order to be sure that effects can be replicated (Mihalic, 2004). In contrast, others explicitly recognised the need for standardising complex interventions by their functions, because these represent interrupting events within the complex systems in which they operate (Hawe et al., 2004a; 2009). Since the precise activities of which an intervention comprises will need to fit the different systems in which it is to be implemented, they argue that to improve the effectiveness of evidence based interventions:

“Rather than defining the components of the intervention as standard [and to be implemented the same at each site] –for example, the information kit, the counselling intervention, the workshops– what should be defined as standard are the steps in the change process that the elements are purporting to facilitate or the key functions that they are meant to have. Intervention integrity would be defined as evidence of fit with the theory or principles of the hypothesised change process”
(Hawe et al., 2004a).

In other words, some authors argue that what might be needed to enhance the successful dissemination and effectiveness of some complex social interventions is a theory of change that explicitly describes the functions that activities at different local sites (during and after trials) should have and on the basis of which implementation can be monitored. Which interventions are suitable for strict standardisation across sites and which interventions are more suitable to be flexible across local sites can only be determined using empirical data. However, the notion that complex social programmes, such as Studio Schools, could be conceptualised and operationalised by their theoretical functions instead of or alongside their form represents an interesting development in intervention research. Some previous efforts have explored the

feasibility and potential utility of this idea. Methods to match behavioural determinants derived from theory to behaviour change techniques to be used in interventions have been developed and widely disseminated (Michie et al., 2008a). Michie & Prestwich (2010) developed a coding scheme for assessing the extent to which interventions are theory-based. This included coding if specified intervention activities are linked explicitly to their theoretical functions, and if the latter are measured as part of the study. Furthermore, the scheme addressed if these theoretical functions were actually monitored at follow-up. It was piloted on 29 primary studies of interventions to increase healthy eating in the general population, and subsequently revised. Their efforts resulted in a 19-item checklist that yielded strong reliability statistics when used by multiple researchers, thus demonstrating that links between intervention components and their theoretical functions can be extracted reliably from reports of primary studies. Similar work was done by Gardner et al. (2010), who used theory to synthesise the effects of audit and feedback interventions that aimed to change professional practice taking place in healthcare settings or targeting healthcare outcomes. In their review, they mapped the intervention activities reported in primary studies to components of a behavioural change theory (Control Theory) in order to analyse through meta-regression whether techniques linked to that theory lead to more effective interventions than techniques without an explicit link to theory. Hardeman and colleagues (2005) developed a causal modelling approach that maps specific intervention techniques to generic ‘intervention points’ (i.e., causal change mechanisms that could be targeted by intervention activities) for use in primary studies. Behaviour change techniques (activities) were selected based on the underlying programme theory, techniques used in other interventions that targeted similar behaviours, and a review of evidence about the effectiveness of these

techniques. Changes in behavioural determinants (i.e. the functions targeted with intervention activities) were measured through the use of questionnaires. The approach developed by Hardeman et al. was subsequently used to categorize specific techniques under common behavioural theories in order to assess the extent to which the latter account for observed changes in outcomes. (Michie et al., 2008b). The efforts described above clearly recognise a distinction between a programme's form and its theoretical underpinnings. However, the doctoral candidate is unaware of any current efforts to explicitly define the nature of a function and its role in programme theory, as well as a method for its operationalisation at the micro-level for the specific purpose of monitoring implementation fidelity. This will be needed not only to explicitly use theory to inform locally designed intervention techniques, monitor their fidelity to theoretical programme model, and improve them, but also to accumulate detailed evidence for and against the theories used to inform intervention programmes.

3.3 Introducing a novel conceptualisation of implementation fidelity

In order to standardise a social programme like Studio Schools by its functions, a good understanding of what constitutes social programmes as well as a redefinition of what constitutes their faithful implementation is needed.

3.3.1 What constitutes social intervention programmes?

A complete definition of the concept of an intervention is offered by Grant (2014), who states that “interventions can be thought of to involve techniques (actions) that aim to modify an otherwise unhindered process or phenomenon (an affair) so as to improve the functioning and well-being of the unit of interest (a course or issue).

Additionally, it is commonly accepted that when studying if and how interventions work, and particularly how to improve them, definitions of intervention programmes also need to include the mechanisms through which actions affect courses or issues. In other words, *the process of intervening* lies at the heart of intervention research. This is recognised by current approaches to intervention development that focus on identifying the malleable risk and protective factors to which specific activities or techniques may be mapped (Bonell et al., 2012; Craig et al., 2008; Fraser et al., 2009), and is also in line with the focus on the *functioning* of programmes that was discussed in section 3.2. Based on the acknowledgement that intervention programmes are not merely actions but also involve the mechanisms targeted by these, the specialist reporting guideline for randomised controlled trials of social and psychological interventions (CONSORT SPI) distinguishes these interventions from biomedical interventions not primarily by the techniques they comprise but instead by the mechanisms of change they target (Grant, 2014). To design interventions with sense and to make them evaluable, they are typically captured in programme theories, leading some to point out they are per definition theories: “programmes are theories incarnate. They begin in the heads of policy architects, pass into the hands of practitioners and, sometimes, into the hearts and minds of programme subjects.” (Pawson & Tilley, 2004). It is argued in this thesis that interventions and theories are technically different concepts, but this definition provides a good explanation of why interventions and the programme theories that describe how they work are frequently considered to be the same for evaluative purposes.

3.3.2 Redefining programme components: activities and functions

Implementation monitoring typically does not focus on a programme ‘as a whole’ but instead targets its specific components (Durlak, 1998; Elliott & Mihalic, 2004).

Therefore, the first step in reconceptualising implementation fidelity to allow for more flexibility in implementation is to disaggregate what constitutes a programme component. In this paragraph, the distinction is highlighted between programme components, activities, and functions, based on the definition of social intervention programmes described in the previous paragraph.

1. Programme activities

Implementation per definition involves activity, as it is through actions that social programmes and interventions cause social change. A *programme activity* is here defined as a specified *activity or set of activities* that is theorised to be related to the effects of an intervention programme and of which the implementation forms part of that intervention programme. Note that programme activities and programme components are currently often treated as more or less the same concept in the literature (Lane et al., 2004; Borelli et al., 2005; Dufrene et al., 2005; Thompson *et al.*, 2003).

2. Programme functions

Programme activities are selected to form part of an intervention programme because of their theorised functions: the system elements, circumstances, cognitive processes etc. they aim to change. A *programme function* is here defined as a *mechanism of change* in a programme's theory of change. This definition of a function is largely borrowed from Hawe et al. (2004a) who defined functions as the 'steps in the change process that [intervention] elements are purporting to facilitate'. However, it is here added that what distinguishes a function from other change mechanisms is that it consists of *the unique and immediate contribution of an implemented activity* or set of activities towards a programme's theory of change, whereas other mechanisms of

change are the direct result of functions or subsequent mechanisms. In other words, where functions occur, implementation ends and the social change process begins. In line with programme theory, a function should lead to a desired output, further change mechanism or outcome, or facilitate the implementation of other activities. This conceptualisation of a function is much more detailed than as presented by Hawe et al. (2004a). It is somewhat in line with the ‘reasoning’ part of a mechanism that was introduced by Dalkin and colleagues (2015), which they defined as the part of a mechanism that represents the volition of stakeholders, as opposed to the ‘resources’ part that represents intervention activities. However, the definition proposed in this thesis not only distinguishes functions from programme activities, but also from other mechanisms of effect in a theory of change, by articulating its importance to implementation fidelity. This has substantial implications that will be discussed in paragraph 3.2.5.

3. Programme components

The programme component is often used as the focus of analyses when connecting implementation data with outcome data. A programme component is here defined as the *combination* of a programme activity or set of activities and its function or set of functions. This definition underlines the fact that components include activities, but are also active in their own right (i.e. fulfil a function), hence the term ‘active ingredients’ (Durlak, 1998) that is often used to describe programme components in the literature. It is also consistent with the observation that mechanisms (in this case functions) are related to but not synonymous with intervention strategies or activities (Jagosh *et al.*, 2011; Lacouture *et al.*, 2015).

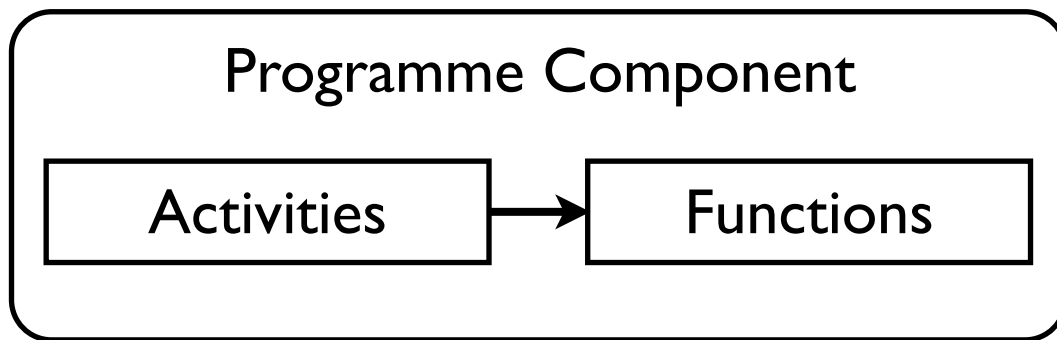


Figure 3.1: A programme component comprises an activity/set of activities and its function(s)

Figure 3.1 represents a programme component in its simplest form as defined in this paragraph. However, the number of activities and functions that are part of any programme component will be determined by that programme's theory of change, and –especially in the case of more complex interventions– activities that form part of one component may also affect functions that form part of other components. The discrete component of which any activity or function is hypothesised to primarily form a part of merely represents the set of activities and functions that were hypothesised to be required for intervention effectiveness.

3.3.3 The relationship between activities, functions, and outcomes

Ultimately, implementation researchers are mainly interested in studying implementation because of its relation to programme outcomes and effectiveness (e.g. Lipsey, 1999; Mihalic, 2004). Determining the relationship between the implementation of programme components and the effectiveness of the programme is also important in order to inform dissemination of evidence-based programmes and the improvement or discard of ineffective services. This paragraph therefore focuses on explaining the assumed relationships between programme components –defined as activities and their functions–, mechanisms of effect, and outcomes.

Establishing the effectiveness of a programme depends on manipulating an independent variable by means of a ‘stimulus’ that in its turn changes a dependent variable or outcome. In its simplest form, ‘implementation’ forms the means by which an independent variable is manipulated. Process evaluations therefore often measure the implementation of programme components, as empirical research has demonstrated that levels of implementation correlate with outcomes within intervention studies (Domitrovich & Greenberg, 2000; Durlak, 1998; Durlak & DuPre, 2008). However, the relationship between the implementation of components and outcomes is often not direct, nor unaffected by other factors. Baron & Kenny’s (1986) landmark work on mediators demonstrated that implementation as a stimulus does not always interact directly with outcomes and that the relationship is often mediated by other variables or may work through effect mechanisms. Weiss (2004) supports this when distinguishing between ‘implementation theory’, which specifies the steps that need to be undertaken during implementation for a programme to be successful, and ‘programmatic theory’, which specifies the theoretical change process that also needs to proceed in line with expectations for the programme to be effective. Pawson & Tilley (1997) state this explicitly:

“Programmes work (have successful ‘outcomes’) only in so far as they introduce the appropriate ideas and opportunities (‘mechanisms’) to groups in the appropriate social and cultural conditions (‘contexts’).”

Discussing the many manners through which contextual factors may influence programme effects, and how to assess such influences is beyond the scope of this thesis. Nevertheless, it is clear that the implementation of activities does not always – likely not in most cases – directly ‘cause’ outcomes.

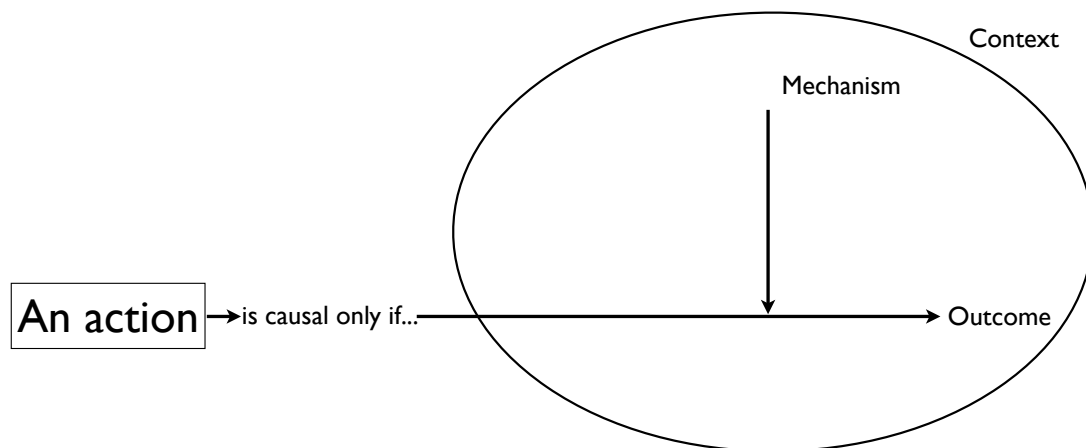


Figure 3.2: An action is causal only if its outcome is triggered by a mechanism acting in context (from Pawson & Tilley, 1997).

The importance of functions within programme design

If the implementation of programme components is the stimulus that triggers mechanisms of effect that cause outcomes to occur, it is important to examine the theoretical implications of distinguishing between activities and functions within the definition of a programme component. In the last paragraph (3.2.1), functions were defined as the direct result of implemented activities, and it was determined that mechanisms of change can only result from functions or other mechanisms of change. From this, it follows that –in theory– the causal impact of activities on mechanisms of change and outcomes should in its entirety be mediated by the functions of the activities. It should be noted that this will only be the case if a theory of change in which the relationships between activities, functions, mechanisms of change, and outcomes are specified is 100% accurate, which will likely never be the case. However, in its pure form, the definition of programme components presented in paragraph 3.2.1 assumes that activities affect mechanisms of change only through their functions. This puts functions at the heart of programme design, as they form the direct targets of programme activities that need not in principle look the same across

contexts as long as they fulfil their function(s). Figure 3.3 graphically presents this idea in its basic form:

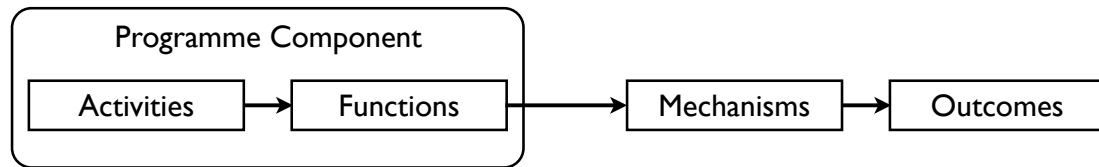


Figure 3.3: Programme activities affect mechanisms of change and outcomes through their functions

In many cases, particularly in the case of complex interventions, social programmes have several components that comprise of several activities and functions. It is also possible that groups of activities may interact to trigger a common function, or that one activity fulfils multiple functions. In addition, activities that are grouped under discrete components (e.g., for the sake of clarity or because they were primarily included to trigger a discrete set of functions) may nevertheless also trigger functions that form part of other components. Figure 3.4 presents a graphic depiction of how the implementation of activities within a more complex social programme could lead to a specified set of functions that in turn could trigger mechanisms that would cause an outcome to occur. As becomes clear from figure 3.4, occurring functions are technically not part of the ‘implementation’ part of a theory of change. This follows from the notion that ‘where functions occur, implementation ends and the social change process begins’ presented in paragraph 3.2.1. Nevertheless, functions are part of programme components, as they are the active mechanism through which components’ activities work.

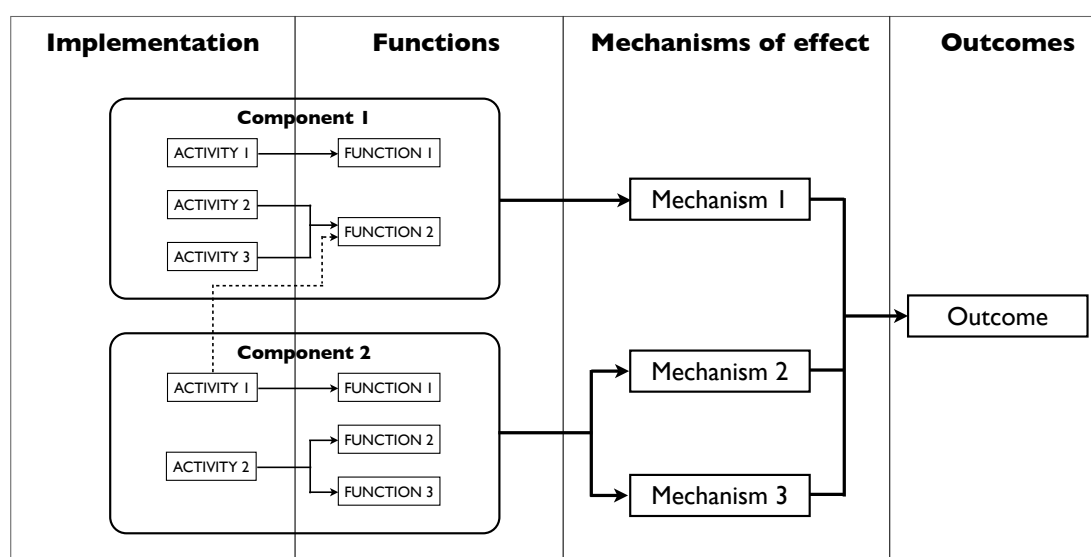


Figure 3.4: Graphic depiction of a hypothetical relationship between programme activities, functions, mechanisms of effect, and outcomes.

3.3.4 Distinguishing between two notions of implementation fidelity

Now that a clear distinction has been made between programme activities and functions and the relationship between these concepts and a programme's change mechanisms and outcomes has been conceptualised, these definitions can be used to inform a new approach towards defining implementation fidelity. Based on the concepts presented in the previous paragraphs, a distinction is now made between a *form-focused* approach and a *function-focused* approach towards defining and monitoring implementation fidelity.

1. Form-focused implementation fidelity

A form-focused definition of implementation fidelity focuses on the implementation of specified programme activities. In order to monitor implementation fidelity based on this approach, activities are specified and subsequently operationalised after which their implementation is monitored. Implementation fidelity is then defined the same as in the 'traditional' and commonly used definition of implementation fidelity: the

degree to which specified procedures or activities are implemented as intended (Dane & Schneider, 1998; Mihalic, 2004).

2. Function-focused implementation fidelity

In contrast with form-focused fidelity, a function-focused definition of implementation fidelity does not focus on activities or procedures but instead on the presence of specified functions. This definition finds its legitimacy in the assumption that programme activities are only relevant for a programme's theory of change because they fulfil a function in it. When monitoring programme fidelity in line with this approach, functions should be specified and operationalised and their presence during implementation should be monitored. Implementation fidelity is then defined as the extent to which specified functions are present during intervention implementation. This definition of implementation fidelity is based on three core assumptions that will need to be explored in empirical studies.

1. Occurring functions are the result of implementation of an activity or set of activities, and the extent to which the presence of specified functions correlates to the implementation of the activity/activities needs to be demonstrated empirically
2. Provided that all functions of all implemented activities are specified appropriately in a valid theory of change, the relationship between implemented activities and outcomes is mediated in its entirety by the functions of the activities
3. The extent to which specified functions are related to specified outputs and outcomes corresponds to the validity of programme theory, and not to the specific form of an implemented set of programme activities

Box 3.1: Three underlying assumptions of a function-focused approach towards defining and monitoring implementation fidelity

3.4 Implications for collection and analysis of implementation data

In order to avoid monitoring the implementation of the Studio School model merely in terms of a set of standardised activities, the presence of the functions of its core

components is also monitored in this thesis. The manner in which implementation fidelity data are used in research studies depends on the purpose for which they are used and analytic decisions made by researchers. Examples include site-level subgroup analysis in which outcomes are compared for ‘high implementation’ and ‘low implementation’ sites (e.g., Battistich et al., 2000), integrating implementation data in a narrative discussion of findings (e.g., Cho et al., 2005), and including implementation as a site-level covariate in prediction models (e.g., Lapan, 2001). However, the methods by which implementation data are *collected* are generally similar across studies. Typically, implementation data come from researcher observations (either in vivo or recorded) or from reports submitted by staff or participants from which ‘fidelity’ scales are constructed for analysis (Breitenstein et al., 2010). The decision to monitor the functions alongside of the activities of PBL and coaching in Studio Schools carries some implications for the type of implementation data that are collected, how these are collected, and how they are incorporated in process evaluations.

3.4.1 Implications for data collection

1. Simultaneous but separate data for activities and functions

Once a theory of change, starting with component functions and ending with outcomes, has been specified and programme activities have been mapped to their functions, data must be collected to monitor the implementation of activities as well as the occurrence of functions. Otherwise, it will be impossible to assess the extent to which implemented activities correlate with their functions at each monitored site. This implies that collecting data on the presence of functions cannot replace monitoring of implemented activities, if a study is to offer any guidance regarding best implementation practices. The assumption that activities and functions are

conceptually different furthermore requires that separate scales can and should be constructed to measure both concepts to facilitate a valid analysis of implementation within process evaluations. This has some implications for the methods that would be used to monitor implementation within research studies.

2. Methods for collecting data on activities and functions

Collecting data on both the implementation of activities and the occurrence of functions likely increases the amount of sources from which data will need to be collected. This is due to the fact that activities and functions are fundamentally different constructs. Because activities take place in physical environments, these can be observed by researchers, staff, and participants (although plenty of obstacles exist when doing this in research studies). Functions, however, form part of a change process (which could consist for example of changing attitudes and beliefs of participants, their learning new concepts and ideas, or even changing social environments in which they interact) that cannot always be observed and may have to be addressed through self report or testing of participants. Therefore, monitoring functions simultaneously to activities could prove to be more resource intensive, and could pose more challenges to reliability and validity, than monitoring activities alone. It needs to be noted here that measuring mechanisms of change alongside implementation efforts is common in process evaluation, and that it is mainly the interpretation of these unique mechanisms that is different from previous efforts.

3.4.2 Analytical implications

1. Assessing the relationship between form-focused and function-focused fidelity data

The purpose of collecting separate implementation data for a programme's form and functions is to assess if the activities that are implemented at study sites correspond to the standardised set of functions these should facilitate. This implies that the

correlation between scores on form-focused and function-focused fidelity scales needs to be determined. The extent of this correlation can then be used to explore the extent to which form-focused and function-focused fidelity co-occur. It needs to be noted here that to determine the *causation* of occurring functions by a set of activities would bring along major implications for research design that are beyond feasible within this doctoral project. These implications will be discussed later on, in paragraph 7.6.1 of Chapter VII.

2. Analysing the pathway from form-focused fidelity to outcomes

Assumption three in Box 3.1 states that the extent to which specified functions are related to specified outputs and outcomes is an indication of the validity of programme theory, and not of successful implementation. Instead, success of implementation depends on the extent to which functions are fulfilled by activities. If implementation is successful but outcomes fail to improve, revision of programme theory should be considered. Once it is determined that form- and function-focused fidelity data are correlated, a programme theory can be tested by analysing the pathway from form-focused fidelity scores to outcomes via function-focused fidelity scores. Although the explicit category labels given to these variables included in such an analysis is new, the analytic technique is not. Mediation analyses –in which a specified variable is hypothesised to mediate the relationship between two or more other variables– are commonly conducted as part of effectiveness studies. Analytical techniques for executing these analyses include hierarchical multiple regression, latent growth modelling, and latent difference score modelling (Dekovic et al., 2012; Gardner et al., 2010; Hogendoorn et al., 2013). Implementation data (typically form-focused) are also included frequently as parameters in outcome analyses (Durlak & DuPre, 2008; Lapan et al., 2001), and it is proposed here that these common

analytical methods could be employed to analyse the pathway from form-focused fidelity to outcomes with a mediating role for functions. Such an analysis was not conducted as part of this doctoral project because implementation data were only collected for two components of Studio Schools during a small sample of sessions, meaning that no valid variance in student outcomes could be predicted based on the data.

3.5 Discussion

3.5.1 Differences with previous efforts to increase flexibility of fidelity

The distinction between form-focused and function-focused fidelity outlined in this chapter is not the first effort to increase flexibility in the monitoring of implementation across sites within research studies. However, it does offer a detailed method for balancing the need to integrate practitioner judgement, client preferences, and constructive adaptation with the need to collect standardised implementation data within rigorous evaluations. Such a method has not been offered within previous efforts to improve monitoring of implementation fidelity in primary studies. Relating implementation fidelity to programme functions was originally proposed by Hawe and colleagues in 2004 (Hawe et al., 2004a). However, because they did not provide a clear and distinctive definition of a function, they could also not define precisely how their notion of fidelity could be used in programme theory development and implementation monitoring in empirical studies. This is illustrated by Table 3.2, which is directly taken from their core paper on the topic. It is noted here that this shortcoming is not unique to Hawe et al., as what is meant by ‘mechanisms of change’ is often poorly defined in the literature (Lacouture et al., 2015). As becomes clear from Table 3.2, ‘standardisation by function’ according to Hawe et al. implies

an understanding of the nature of functions that is different from the one presented in this chapter. Instead of standardising the change mechanisms that form the direct result of implemented activities, the types of standardisation presented in Table 3.2 in fact both describe *activities*.

*Table 3.2: Examples of standardisation by form and function
(as defined by Hawe et al., 2004a)*

Principle of intervention	Type of Standardisation	
	By form	By function
To educate patients about depression	All sites distribute the same written patient information kit	All sites devise ways to distribute information tailored to local literacy, language, culture, and learning styles
To improve detection, management, and referral of patients in primary care	All sites hold a series of three in-service training workshops for general practitioners with preset curriculums	Local health authorities are provided with materials and resources to devise in-service training tailored to local schedules, venues, and preferred learning methods
To involve local residents and decision makers in order to increase uptake, effectiveness and sustainability of the intervention	A local intervention steering group committee is convened in each site with representatives of pre-specified organisations	Mechanisms are devised to engage local key agencies and consumers in decision making about the intervention. Suggested options: steering committee, consultations, surveys, website, phone-ins
To harness and facilitate material, emotional, and affirmational support across social networks of people in particular life stages	All mothers of new babies are invited to join discussion and mutual support groups. People moving into nursing homes receive three friendly visits from a designated resident	Methods to alter network size, network diversity, contact frequency, reciprocity, or types of exchanges are tailored to subgroup preferences

The difference between standardisation by form and by function in this table is merely the level of detail in which activities are standardised. If we consider, for example, the first principle described in the table: educating patients about depression. The suggested standardisation by form is a standardised activity in the form of distributing an information kit. The suggested standardisation by function is also an activity (distributing information): the difference with standardisation form is in the level of standardisation of the activity. This is in line with the approach towards flexible activity monitoring that is proposed in this chapter, but does not provide any guidance

as to how to standardise the functions these activities should have. In other words: the assumption that different activities fulfil the same function cannot be verified empirically. Other authors have also allowed or argued for large amounts of flexibility in the interpretation of programme activities by local implementers (for example: Battistich et al., 2000; Saunders et al., 2006; McHugh et al., 2009) and there are considerable merits in flexibly defining programme activities. However, doing so without explicitly connecting the activities –theoretically and empirically– to specified functions complicates the accumulation of research evidence for programme theories and specific combinations of activities (Michie et al., 2008a). In addition, it makes it difficult to advise practitioners regarding best practices and to hold researchers and implementers accountable to the results of their efforts. The approach presented in this thesis offers a methodological framework –the first to the knowledge of the DPhil candidate– to link locally defined activities to standardised functions and to measure the adequacy with which this is done.

3.5.2 The utility of this approach for implementation and evaluation research

Distinguishing between form and function when monitoring implementation fidelity offers some new opportunities to increase the evidence base for social intervention programmes and theories, as well as improve the uptake of evidence-based services. Furthermore, it does so without ignoring the contextual variation in implementation settings presented by a complex world that are often presented as a major threat to, or even an argument against, evidence-based practice.

Alleviating the tension between fidelity and adaptation

Standardisation of programme components lies at the heart of the tension between implementation fidelity and contextual adaptation of complex social interventions. It

is argued here that this tension occurs when programme activities –or a programme’s form–, are standardised as an intervention ‘package’ whilst circumstances during implementation require adaptation of this package. In paragraph 3.3.3, the relationship between functions and activities was explained as one in which the former affect programme effects exclusively through the latter, thus putting functions at the heart of programme theory. A function-focused approach towards implementation fidelity would not remove standardisation as part of programme development or evaluation, but would focus it on a programme’s functions instead of its form. Combinations of activities could be specified locally and mapped to functions (Bonell et al., 2012; Hawe et al., 2004a). Although implementation of activities will still need to be monitored (which can be an extensive task if specified differently for each site of implementation) it will not need to be the same set of activities in every site to preserve a standard programme theory that can be monitored in line with best practices in intervention research. This provides one way of alleviating the tension between preserved implementation fidelity and contextual adaptation.

A more systematic approach towards improving programme theories

Theory-driven evaluation focuses not only on the question if interventions work, but also on how they work and for whom (Chen & Rossi, 1983; Weiss, 1997). Realist evaluators further stress the importance of contextual factors, so that evaluation comes to focus on ‘what works for whom in what contexts’ (Pawson, 2013). This focus on the process of interventions –as opposed to just their outcomes– intends to fill in the ‘black box’ of effectiveness studies, and is now common practice in evaluation research (Fraser et al., 2009). The core implication of such approaches towards evaluation is specificity. As previously argued in Chapter II, programme

theories need to be specified explicitly and in detail in order to be tested rigorously and subsequently improved. Furthermore, it is recognised that approaches for determining how contextual factors should affect programme theories and their implementation through intervention programmes need to be developed and operationalised prior to commencement of evaluation (Bonell et al., 2012; Pawson & Tilley, 1997; Weiss, 1995). However, the literature lacks systematicity when it comes to the process of evaluating and revising programme theories based on the findings of process evaluations. One important question here is: *what in a programme theory needs to be revised for a particular context as a result of how it is influenced by that context?* This question is core to the successful dissemination of evidence-based practices, and authors have argued that it has been insufficiently addressed (Bonell et al., 2006; Cartwright & Hardie, 2012). Distinguishing between activities, functions, and further mechanisms of effect within an intervention process that works towards outcomes allows for a more systematic approach towards answering this question. Table 3.3 provides an overview of how programme theories could be modified based on the findings of process evaluations or could be adapted proactively to a specific context based on prior research or experience in a systematic manner. In essence, a programme theory could be modified in terms of its activities or its functions or both, or the appropriateness of the theory for a given problem in a given context could be reconsidered altogether.

Consider a hypothetical example of the implementation in UK secondary schools of a smoking prevention intervention that was successfully tested and disseminated in the United States. A programme theory comprising standardised activities, functions, further change mechanisms, and outcomes would have been evaluated repeatedly with success in different US settings.

Table 3.3: A systematic approach towards modifying programme theories based on anticipated or demonstrated contextual influences

Anticipated/Demonstrated contextual influence	Response
The relationship between form and functions is (expected to be) affected by a specific context, but the relationship between functions and change mechanisms/outcomes is (expected to be) left intact	Programme theory can be left intact, but implementation should be revised by selecting activities appropriate for the specific context
The relationship between form and functions is (expected to be) left intact within a specific context, but the relationship between functions and change mechanisms/outcomes is (expected to be) affected	Programme theory should be revised by removing or adding relevant functions (in)appropriate for the specific context, but implementation of those activities that trigger relevant functions can be left intact
The relationship between form and functions, as well as the relationship between functions and change mechanisms/outcomes are both (expected) to be affected by a specific context	The appropriateness of a programme theory for intervening with a problem within the specific context should be reconsidered

During piloting, UK evaluators may find that the core functions that produced important change mechanisms in the US (e.g., attitudes towards smoking) worked essentially the same in the UK (i.e., show similar correlation to outcomes), but that not all activities (e.g., presenting statistics on the association between smoking and health outcomes) triggered these functions as predicted by US results, regardless of their implementation with fidelity. This problem could then be labelled as an implementation problem instead of a theoretical problem and activities could be modified accordingly whilst functions could be left intact. Of course, alternatives to this example are possible, including simultaneous revision of activities and functions as well as rejection of the programme for the UK context altogether. Nevertheless, the distinction between the form, functions, and subsequent change process within programme theories allows for their systematic evaluation and improvement.

3.6 Chapter summary

In this chapter, the approach according to which implementation fidelity is defined within this thesis was introduced. Because of the context of the project, an evaluation

of implementation practices of a social programme that is highly susceptible to adaptations, implementation fidelity had to be reconceptualised. It was done so according to a distinction between form- and function-focused implementation fidelity, which was based heavily on previous work conducted by Hawe and colleagues (2004a). However, the core concepts that form part of the distinction were defined clearer and its analytical implications as well as implications for implementation and evaluation research were discussed in much further detail. In the remainder of this thesis, both concepts will be revisited repeatedly as the methodology and results of the project are reported and discussed. First, the next chapter will present how implementation measures were developed on the basis of the concepts described as part of the current chapter.

Chapter IV: Defining Implementation Measures for Studio Schools

4.1 Introduction

The previous chapter explained the conceptual difference between form-focused and function-focused implementation fidelity monitoring, and provided the motivation for monitoring both kinds of fidelity in Studio Schools. In the current chapter, the process of specification and subsequent operationalisation of implementation measures for two components of the Studio School model is described.

4.1.1 Focus on project-based learning and personal coaching

In order to facilitate a feasible in-depth analysis of the implementation practices of different Studio Schools, it was agreed with the Studio Schools Trust (SST) to monitor the two essential school-based model components: i) project-based learning (PBL) and ii) the personal coaching of students (see background on Studio Schools). Monitoring implementation of the third active component of the model, external work placements, is considered to be beyond feasibility giving the limited resources available for this project. Additionally, the SST and individual Studio Schools expressed concerns with engaging employers in the study as partnerships have only recently developed and participation would have imposed a substantial burden on employer partners.

4.1.2 Fidelity as implementation dimension of focus

The measures described in the current and further chapters only measure one dimension of programme implementation: fidelity. It is acknowledged here that there are other important dimensions of implementation that are not measured in this DPhil thesis, including dosage, quality, uptake, and participant responsiveness (Dane &

Schneider, 1998). However, as demonstrated by the previous chapter, it was determined that limited resources could be put to best use towards advancing the literature by focusing on a reconceptualisation of the concept of fidelity. Hence, the project focuses on this dimension of implementation. Further reflections on this decision are discussed in paragraph 6.4.3 of Chapter VI.

4.1.3 Overview of methods and procedures

The measures used in this study were developed and refined during a multi-stage process. In the current chapter, the process of developing measures –as summarised in Figure 4.1– is reported, whereas the next chapter addresses the analytical procedures used to refine these upon completion of data collection. Each step in Figure 4.1 will be discussed in detailed in this chapter, starting with the development of detailed theoretical frameworks in the next section (4.2). A brief description of each step, along with used methodology, is provided here for the purpose of clarity.

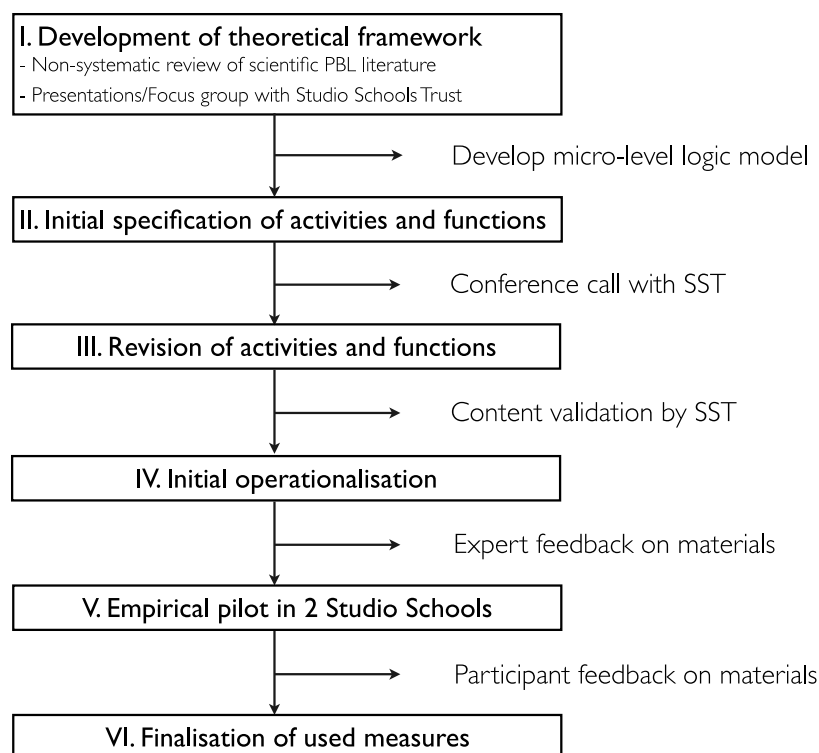


Figure 4.1: Flowchart depicting the development of implementation measures for Studio Schools

First, detailed theoretical frameworks that describe distinct programme components were specified for PBL and coaching in Studio Schools, based on inspection of the scientific literature and interactions in the form of presentations hosted by and a focus group with members of the Studio Schools Trust (SST). The procedure for developing these frameworks as well as a set of logic models in which they are summarised are discussed in detail in section 4.2. Next, for each component, activities and functions were specified based on the same sources that informed the original theoretical frameworks. The lists of activities and functions were revised based on a conference call with the SST, after which they were finalised and approved for their content validity. This process is reported in section 4.3. All activities and functions were then operationalised as quantitative questionnaire items, to be rated by researchers, teachers, and students on ordinal-level Likert-scales. The questionnaires were given to fellow researchers with expertise in education and psychometric assessment, as well as teachers and students to gain feedback regarding their quality in terms of ease of use and comprehensiveness. The questionnaires were revised upon feedback reception and were piloted in two Studio Schools prior to their finalisation. The process of expert- and participant consultation as well as piloting is described in section 4.4. Section 4.4 also presents the measures of implementation as used within the study. A discussion of the methods and procedures that are reported in this chapter is provided in section 4.5.

4.2 Basic theoretical framework

Any set of rigorous measures employed to monitor implementation should be rooted in an explicit theoretical framework (Cook et al., 2000; Saunders et al., 2006; Weiss, 1995). Between October 2013 and February 2014, the basic micro-level theoretical

frameworks for PBL and personal coaching were developed. These were then revised during the iterative instrument development phase that is presented in paragraph 4.4.2 until a final framework emerged.

4.2.1 Procedure

The basic theoretical frameworks for PBL and personal coaching in Studio Schools were developed during a three-staged process. First, information was collected from various sources to inform original drafts of the frameworks. These sources included a non-systematic ('snowball') review of the scientific literature, attendance of presentations given by the Studio Schools Trust on the specific models of PBL and coaching recommended for Studio Schools, and a one-hour focus group interview with members of the Studio Schools Trust. During this focus group, SST members were asked to specify the most important components of PBL and coaching. Second, the input gathered from the sources specified above was used to distil components of PBL and coaching and map these to the elements of the Studio Schools Logic Model (see Chapter II) that the school-based curriculum feeds into. Micro-level logic models were then drafted for both components. Third, these logic models were revised based on feedback received from the SST during a conference call in which the concrete activities and functions of PBL and coaching were discussed (see paragraph 4.3.1). The basic theoretical frameworks that resulted from this process are described and summarised in Figures 4.2 and 4.3.

4.2.2 A basic theoretical framework for project-based learning in Studio Schools

The basic theoretical framework used to guide the development of implementation measures for PBL in Studio Schools is presented succinctly below. Box 4.1 presents a brief overview and description of the six PBL components that emerged during the

theory development process. Figure 4.2 depicts the theorised relationships between components, mechanisms of effect, and outcomes. For a more detailed discussion of how the PBL literature formed one of the sources on which this framework was based, see Section 1.3.2 in Chapter I.

A. Learning and developing skills as a means to an authentic end

Students learn subject content (e.g. English, Maths) and develop skills as a means to solving a presented problem that could be encountered in the world outside school.

B. Engaging in a challenging enquiry process

Students independently find solutions to the problems they encounter and self-identify what subject content and other information they need to find and master to complete their projects.

C. Obtaining and applying subject knowledge

Towards completing their projects, students study and apply content from the core subjects offered at their Studio School and understand how content/knowledge used in projects aligns with content/knowledge required for assessments (i.e. exams).

D. Explicit development of CREATE®-skills (non-cognitive skills)

During the process of working on their projects, students reflect explicitly on their CREATE-skills.

E. Elaborating on subject content and reflecting on own work

Students submit multiple drafts of their project work and receive, as well as provide, feedback on their and others' work.

F. Application and assessment of knowledge and skills in multiple ways and contexts

Students learn how applied subject content can be of use in contexts beyond school.

Box 4.1: Components of project-based learning as theorised for Studio Schools

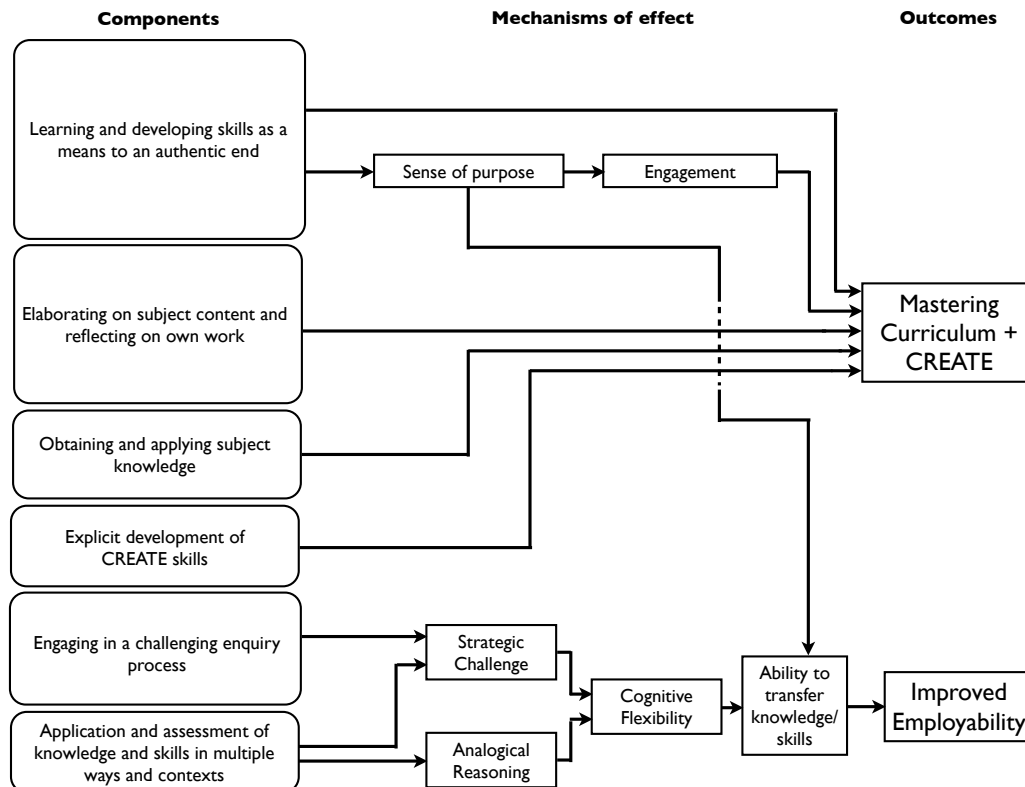


Figure 4.2: A basic overview of the components, mechanisms of effect, and outcomes of project-based learning as theorised for Studio Schools

4.2.3 A basic theoretical framework for personal coaching in Studio Schools

The basic theoretical framework used to guide the development of implementation measures for personal coaching in Studio Schools is presented succinctly below. Box 4.2 presents a brief overview and description of the four personal coaching components that emerged during the theory development process. Figure 4.3 depicts the theorised relationships between components, mechanisms of effect, and outcomes. For a more detailed discussion of how the coaching literature formed one of the sources on which this framework was based, see Section 1.3.2 in Chapter I.

A. Guided self-reflection on performance and behaviour

Guided by their coach, students become aware of the perceptions and thought patterns that influence their strengths and weaknesses.

B. Joint setting of goals/solutions and planning/reflecting on actions

Students and coaches jointly specify the goals students work towards and what is needed to achieve these.

C. Self-direction towards goals and solutions

During coaching sessions, changes needed to improve students' situations (e.g. behaviour, performance) are specified primarily by the students.

D. Supportive relationships

Students develop supportive relationships with adult individuals that are aware of and understand their personal situation and characteristics.

Box 4.2: Components of personal coaching as theorised for Studio Schools

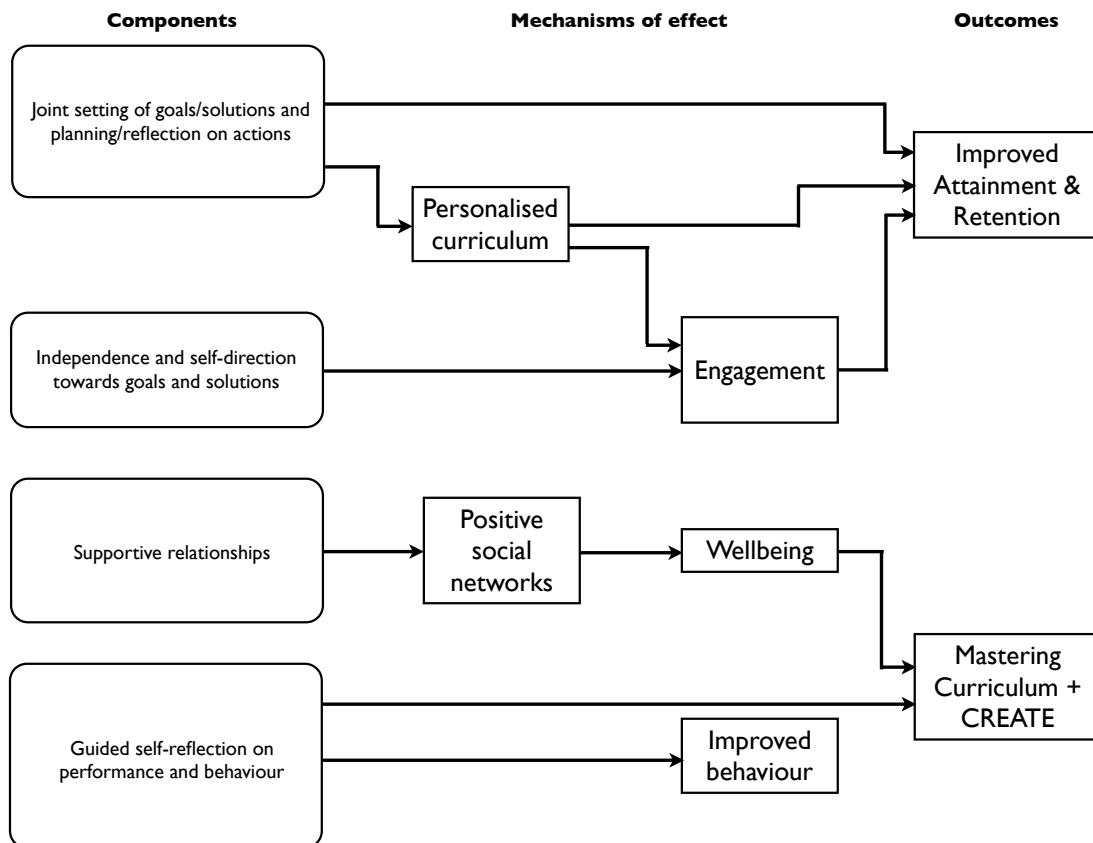


Figure 4.3: A basic overview of the components, mechanisms of effect, and outcomes of personal coaching as theorised for Studio Schools

4.3 Specifying the activities and functions of PBL and coaching

In paragraph 3.3.2 of Chapter III, programme components were defined as the combination of a programme activity or set of activities and its function or set of functions. After the components of PBL and personal coaching were defined, the next step in developing implementation measures for Studio Schools entailed defining the functions and activities of which these components comprised. This process was completed between February and March 2014.

4.3.1 Procedure

The same sources (non-systematic literature review, presentations by the Studio Schools Trust, and a focus-group interview) that informed the previously described theoretical frameworks of PBL and coaching were also used to inform a first specification of activities and functions for each PBL and coaching component. In this original specification, each of the six PBL and four coaching components was divided into a set of numbered unique items that were mentioned as key to PBL or coaching in at least one of the used sources. Some of these items classified best as activities –these implied a specific action– whereas others classified best as functions: these implied a change mechanism resulting from an activity. Second, these items were discussed with four members of the Studio Schools Trust responsible for the dissemination of PBL and coaching in a conference call held in March 2014. During the conference call, SST members suggested the rephrasing of some components and items and also suggested one additional PBL component (D: Explicit development of CREATE®-skills) as well as some additional items. Feedback from the conference call was used to draft a second list of items that was subsequently approved by the SST. For each item that classified as an activity, functions that fit the theoretical

frameworks specified in paragraph 4.2.2 and 4.2.3 were then formulated, and vice versa. At this point, the decision was made to standardise not only the functions but also the activities of the Studio School model for monitoring in this study. The motivation behind this decision was dual. Its primary aim was to assess deeper the dynamics between standardised activities and functions, so as to find how fluctuations in the former related to differences in the latter. It was also judged unfeasible within a doctoral project with limited funding to specify local activities with each individual school, at the same time during which the definitions, logic, and approach supporting a function-focused approach to fidelity monitoring had to be developed from scratch. Third, the resulting list of items was operationalised into checklists and piloted in two Studio Schools (see paragraph 4.4.2). During this piloting process, some items were rephrased or (dis)integrated to better suit the context of Studio Schools and facilitate a rigorous process of data collection. Because most items were rooted in theories developed beyond the setting of secondary schools (e.g., in higher education or in business) some had to be rephrased somewhat to be appropriate for Studio Schools. For example, the coaching of students towards improved educational outcomes was rooted in the principles of coaching in management (in which employees or managers are coached towards improved performance) and had to be tailored to the context of secondary education. In some cases, items were disintegrated into multiple unique items to avoid ambiguity in the questions asked to staff and students, and researchers. Finally, upon completion of the piloting process, the lists of activities and functions of PBL and coaching that were used for collecting implementation data in Studio Schools were created. These lists are presented and discussed below.

4.3.2 Activities and functions of PBL and coaching in Studio Schools

A total of 14 activities and 16 functions were defined to constitute the six components of PBL in Studio Schools. For personal coaching, a total of 10 activities and an equal number of functions were defined for four components. Tables 4.1 and 4.2 present a description of the activities and functions for each component. As becomes clear from this tables, the number of activities and functions for each component is often not the same. Figures 4.4 and 4.5 depict the integration of the PBL and personal coaching activities and functions within their micro-level theories of change. Note that all functions in the logic models are mapped quite linearly to their corresponding activities. This has been done for the purpose of explicitly demonstrating how activities bring about functions that form the beginning of a theory of change.

*Table 4.1: Overview of the activities and functions of each component of **project-based learning** as theorised for Studio Schools*

Activities/Project design features	Functions
A. Learning and developing skills as a means to an authentic end	
AA1. The project should have at least one concrete end product that will be exhibited publicly to an audience beyond the school AA2. The problems or assignments central to project activities should be presented without structured guidance on where to start in order to complete it AA3. One ore more adults from outside the school with expertise/experience that is of relevance to the project should be involved in co-designing or commissioning the project AA4. One or more adults with expertise/experience that is of relevance to the project should be involved in providing instructional support and assessment towards project activities AA5. During the process of working on the project, students should work or meet with at least one adult with expertise that is of relevance to the project	FA1. It should be clear to students and learning coach what the concrete product(s) of the project is/are FA2. The problem central to the project should be of importance to students FA3. It should be clear to students and learning coach how studied subject content aligns with the end product(s) of the project FA4. Students' application of knowledge and skills reflects that of adult life FA5. The adults involved in the project should inspire students to put in their best effort

B. Engaging in a challenging enquiry process	
AB1. When assisting students, learning coaches should not present solutions to encountered problems AB2. When assisting students, learning coaches should not instruct students what they need to study to complete the project successfully AB3. Students should draw upon key methods of enquiry suitable with the project's discipline(s) (e.g., document analysis for English/History, collecting own data for Science)	FB1. Students should discover independently where/how the subject content and other factual information that is needed to understand/manage a problem and complete the project can be obtained FB2. Students should feel ownership of the project and its end product(s): it is their responsibility to come with the answers to complete it successfully
C. Obtaining and applying subject knowledge	
AC1. Students should draw on content of one or multiple subject(s) that are offered at Studio Schools (i.e. English language, English literature, Maths, Science) whilst completing the project.	FC1. It should be clear to students and teachers how content/knowledge that is needed to complete the project links to the subject content they require to learn to pass their exams FC2. Over the scope of the project period, students should have covered an amount of National Curriculum similar to what they would have covered had they attended regular Academies
D. Explicit development of CREATE®-skills	
AD1. During the process of working on the project, students should reflect explicitly on which of their CREATE skills they need to further develop to successfully complete the project	FD1. It should be clear to a student and learning coach which CREATE skills a student needs to develop further to complete the project successfully FD2. It should be clear to a student and learning coach how students' project activities help them develop <u>specific</u> CREATE skills
E. Elaborating on subject content and reflecting on own work	
AE1. Students should receive, as well as provide, explicit oral written feedback/critique on the products (deliverables) of their project activities, including the project's end product(s) AE2. Students should submit multiple drafts of the end product of the project and should explicitly reflect on each draft guided by feedback of the teacher/learning coach AE3. Classroom time should be spend as much as possible on working with subject content (independently developing/applying knowledge) and as little as possible on lecturing subject content	FE1. Students should have sufficient opportunities to express thoughts and ideas about subject content to others (teacher/learning coach and fellow students) FE2. Students recognise those aspects of their own and others' work that can/should be improved FE3. Students recognise those aspects of their own and others' work that are particularly strong FE4. Students develop their own best way of learning/studying subject content for it to be of use towards project tasks and end products

F. Application/assessment of knowledge and skills in multiple ways and contexts	
AF1. During the process of working on the project, students should submit work and be assessed through a range of exercises.	FF1. It should be clear to students how knowledge of subject content attained in the context of the project could have utility in other contexts

Although assumption 2 in Box 3.1 –which states that all effects on outcomes by activities are entirely mediated by functions– requires each activity to fulfil at least one function, it is unlikely that the relationship between activities and functions within programme components is strictly linear. Rather, it may be expected that within complex interventions, activities influence functions through a more synergistic process in which the sum (functions) of a set of activities forms more than its parts (Bonell *et al.*, 2012). To assess exactly the complexity (or simplicity) of the relationship between the activities and functions of PBL and coaching in Studio Schools requires empirical data.

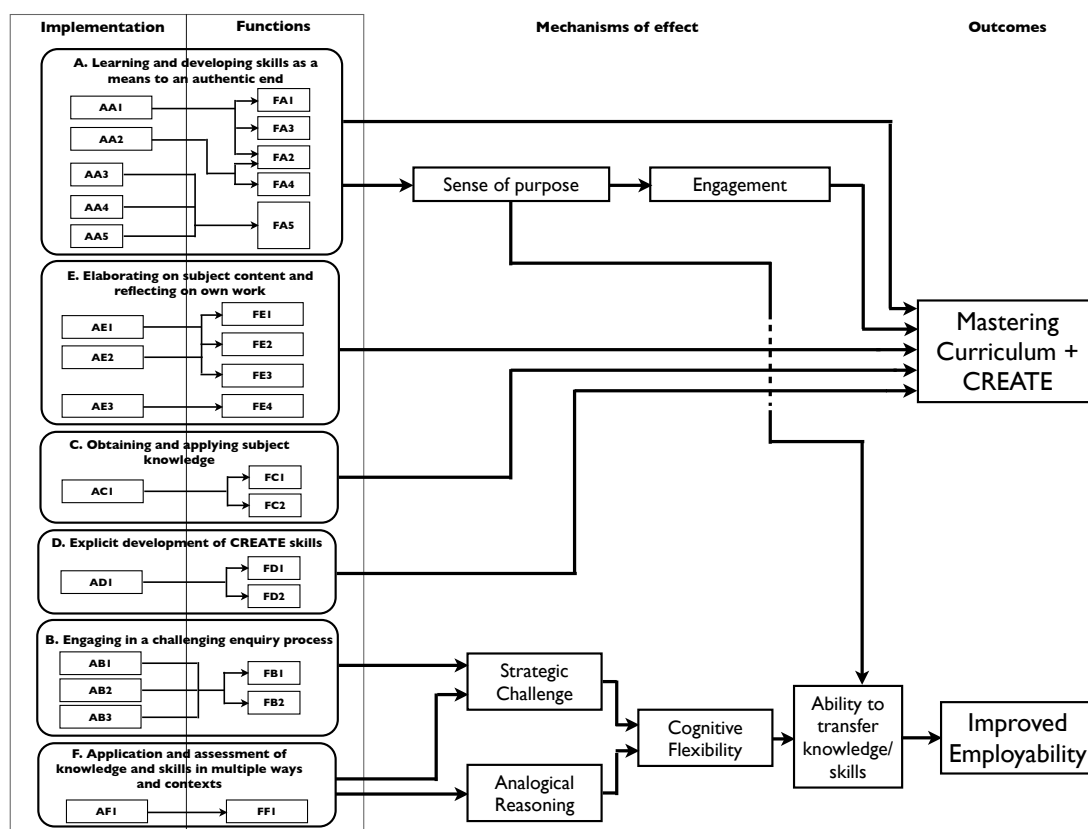


Figure 4.4: A logic model depicting the relationships between activities, functions, mechanisms of effect, and outcomes of project-based learning as theorised for Studio Schools

Table 4.2: Overview of the activities and functions of each component of personal coaching as theorised for Studio Schools

Activities	Functions
A. Guided self-reflection on performance and behaviour	
AA1. Student and coach should reflect on student's academic and CREATE progress during coaching sessions AA2. Student and coach should reflect on student's behaviour AA3. Student and coach should discuss the cues and situations that positively or negatively affect student's performance at school/work AA4. During coaching sessions, a coach should guide the student to discovering methods for reinforcing constructive thought patterns and emotions and reshaping dysfunctional ones.	FA1. After a coaching session, students and coaches should be aware of students' specific strengths and weaknesses with regards to their academic and CREATE progress. FA2. After a coaching session, students and coaches should be aware of students' specific strengths and weaknesses with regards to their behaviour FA3. After a coaching session, students and coaches should be aware of the thought patterns that influence students' perceptions of themselves and others positively and negatively. FA4. After a coaching session, students should be aware of specific methods for reinforcing constructive thought patterns and emotions and reshaping dysfunctional ones.
B. Joint setting of goals/solutions and planning/reflecting on actions	
AB1. During coaching sessions, student and coach should discuss and determine concrete curricular and personal goals of the student. AB2. During coaching sessions, student and coach should discuss and determine the concrete actions that need to be undertaken to achieve these goals AB3. Student and coach should discuss the execution of planned actions to determine progress towards achieving the student's goals and (if necessary) to plan new actions	FB1. It should be clear to the student/personal coach and rating researcher what the curricular and personal goals are that the student is working towards. FB2. It should be clear to students/personal coach and rating researcher what actions are to be undertaken by i) the student and ii) the coach to achieve these goals.
C. Self-direction towards goals and solutions	
AC1. During coaching sessions, coaches should use questions to help students self-examine their behaviours and attitudes. AC2. During coaching sessions, coaches should not offer concrete advice or direct students towards solutions or actions.	FC1. After each coaching session, student/coach and rating researchers should feel that specified changes that are needed to improve student's behaviours/situation/learning were defined primarily by the student.
D. Supportive relationships	
AD1. During coaching sessions, coach should provide sufficient opportunities for students to discuss personal matters (such as experiences at home, at school, and at work) without feeling exposed.	FD1. Students and personal coaches should have a relationship based on mutual trust FD2. Students and personal coaches should feel that the coach is aware of and understands students' personal characteristics FD3. Students should feel supported by their personal coach when engaging with new activities or (interactive) strategies with which they may be uncomfortable.

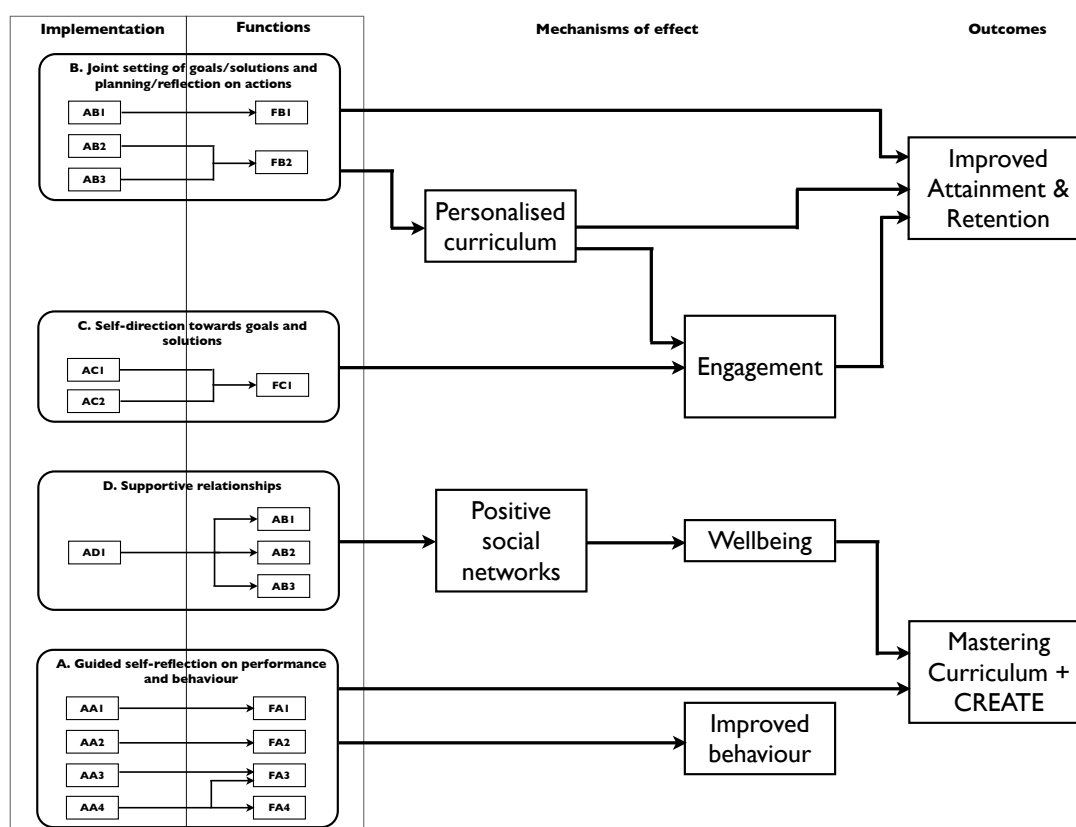


Figure 4.5: A logic model depicting the relationships between activities, functions, mechanisms of effect, and outcomes of personal coaching as theorised for Studio Schools

4.4 Operationalising the activities and functions of PBL and coaching

Little empirical work exists on systematic and comprehensive strategies for measuring implementation fidelity (Breitenstein et al., 2010). Typically, quantitative implementation scores are obtained through questionnaires or checklists that are administered by independent observers (i.e., researchers) or practitioners (Durlak & DuPre, 2008; Forgatch et al., 2006; Kam et al., 2003). In some cases, ratings by programme participants are included (Cho et al., 2005), and qualitative implementation data are also commonly collected (Dusenbury et al., 2005). Fidelity ratings and qualitative analyses can be based on in-vivo observations or recordings with video and/or audio equipment. In this study, quantitative and qualitative implementation data were collected through questionnaires completed by various rater

groups. These rater groups, as well as the procedure for developing the questionnaires, are discussed in this section.

4.4.1 Data sources used

In line with common practice in process evaluation, it was decided to measure implementation fidelity (form-focused and function-focused) quantitatively in this DPhil project through multi-item scales rated in questionnaires. To facilitate triangulation, the decision was made to obtain fidelity data from three sources: the researcher, teachers, and students. An example of how a specific item was presented to each rater group is presented later in this section in Table 4.8.

Researcher checklists

Because independent researchers do not play an active part in the implementation process and are expected to enter fidelity monitoring from a position of equipoise, fidelity data resulting from the ratings of researchers are seen as relatively objective (Breitenstein et al., 2010; Dusenbury, 2003). Therefore, the primary source of PBL and personal coaching fidelity data used on this project comprises quantitative ratings of form- and function-focused fidelity items by the DPhil candidate (FvU). One part of these ratings is based on observations of PBL and personal coaching, and another part on qualitative data provided by teachers and students (see ‘Teacher questionnaires’ and ‘Student questionnaires’). Personal coaching sessions were video-recorded or audio-recorded if participants provided consent for this, so that sessions were minimally disturbed and independent ratings by additional researchers could be facilitated without the need to travel to schools. Initially, it was considered to also video- or audio-record PBL sessions. However, it became clear during pilots that recording PBL sessions would be unfeasible due to 1) an inability to obtain informed

consent from all involved students, and 2) the fact that students in some cases had to relocate across the schools to work on their projects. This resulted in the decision to observe PBL sessions in-vivo.

Staff questionnaires

Implementation data are often also collected from practitioners because of the benefits of triangulation and because their reports may provide important information regarding the portability of interventions during their dissemination (Battistich et al., 2000; Breiteinstein et al., 2010; Cook et al., 2000; Saunders et al., 2006). Because Studio School teachers and coaches are the implementers of PBL personal coaching in Studio Schools, fidelity data collected from staff are also included. Facilitating teachers and coaches were asked to complete one questionnaire for each observed PBL or coaching session. The teacher questionnaires comprise of quantitative and qualitative parts. The quantitative parts contain a set of checklist items similar to those rated by the researcher and facilitate the assessment of reliability of ratings across observers from different perspectives. The qualitative parts contain a series of questions that provide the data used by the researcher to rate those items of the function-focused fidelity scales that cannot be observed in-vivo. As indicated in paragraph 3.4.1 of Chapter III functions form part of a change process in participants' experiences that cannot always be physically observed. Therefore, it was decided that some additional exercises in the form of answering open-ended qualitative questions about their experiences during implementation would have to be obtained from teachers. Their answers would subsequently facilitate researchers' ratings of function-focused fidelity. For example: teachers were asked to report on how they believed the knowledge obtained by students during their projects could be useful in other contexts, so the researcher could assess if a teacher was indeed aware of this function.

Student questionnaires

It is recognised that implementation data collected from programme participants can provide valuable insights that data from independent observers and practitioners alone cannot (Breitenstein et al., 2010; Cho et al., 2005). Therefore, in addition to staff members, students were also asked to complete one questionnaire for each observed PBL and coaching session in which they took part. The student questionnaires contain quantitative and qualitative parts that serve the same purposes as those included in the teacher questionnaires. Apart from a few unique student-rated or staff-rated items in the PBL questionnaires, the student questionnaires are organised identical to the staff questionnaire as to optimally facilitate assessments of reliability between raters.

4.4.2 Developing checklists and questionnaires

First drafts

Once a list of the activities and functions of the components of PBL and coaching had emerged (see paragraph 4.3.2) and the sources from which fidelity data were to be collected were determined, first drafts of the researcher checklists and staff-and student questionnaires were developed. First, it was determined which rater groups could appropriately provide data on the implementation of each activity and presence of each function. In most cases, it was judged that an activity or function could be rated validly by each rater group. For example, for PBL activity ‘AA5’ –relating to students having worked or met with an external adult as part of their projects– or personal coaching function ‘FA3’ –relating to awareness by student and coach of students’ thought patterns that influence their perceptions– it was determined that these could be rated by researchers as well as teachers, and students. In some cases, it was judged that an activity or function could only be rated validly by one or two of the involved data sources. For example, for PBL function ‘FB3’ –relating to students

feeling ownership of their projects– it was determined that a rating provided by a researcher would not yield a valid estimate of its underlying construct. The second step in the process involved assessing which activities and functions could not be rated by researchers based on in-vivo observations or recordings of sessions alone. For these activities and functions, it was decided that additional qualitative answers provided by staff and students were needed. Finally, the first drafts of the researcher checklists as well as the staff- and student questionnaires were created.

Subsequent drafts

The first drafts of the checklist and questionnaires were sent to the doctoral supervisors and a fellow graduate student with teaching experience in secondary schools for feedback on content and phrasing of questions. Their feedback was incorporated after which second versions were drafted. For example, if colleagues indicated they suspected students would not easily comprehend one or several items, suggestions for rephrasing were obtained and incorporated in revised drafts. For the student questionnaire, Flesch-readability indices were calculated after which the questionnaire was rephrased until these were deemed acceptable. All second drafts were sent to the doctoral supervisors again as well as to two academic staff of the Department of Social Policy and Intervention (Oxford) that formed the Transfer of Status panel in order to assess if their suggestions expressed during the ToS process had been incorporated appropriately. After the second iteration, the researcher checklist was deemed appropriate for piloting in Studio Schools. The teacher and student questionnaires were revised once more based on received feedback from the sources above.

Piloting

- Researcher checklists

The PBL researcher checklist was piloted during four visits to two Studio Schools during June and July 2014, in which a total of five PBL sessions were scored using the checklist. During three of these sessions, the checklist was completed by an additional graduate student to assess reliability and discuss methods to improve it. The personal coaching researcher checklist was piloted on recordings of four coaching sessions made in June and July 2014 in two Studio Schools. The checklist was completed once by an additional graduate student. During the piloting process, some items were rephrased or disaggregated into separate items to remove ambiguity. In addition, it was decided to develop fixed scoring guidelines for both checklist to increase reliability of ratings. Final versions of the checklist and the scoring guidelines were drafted in July 2014.

- Staff and student questionnaires

Staff and student PBL questionnaires were piloted during four visits to two Studio Schools during June and July 2014, in which a total of four PBL sessions were scored by staff members and students. Personal coaching questionnaires were piloted during the same visits: six sessions were scored by the participating staff member and student. Participants were asked to indicate on their questionnaires if they felt a question was difficult to understand or otherwise in need of improvement. In addition, oral feedback on the questionnaires was obtained from other staff members during the visits. During the process, the questionnaires were revised based on staffs' and students' answers and comments, as well as the researcher's own experiences. Specifically, for the PBL questionnaire, it was determined to reallocate specific activities and functions to a separate questionnaire that would only be asked at the

onset of a project because these involved structural elements instead of items representing a process that requires repetition in lessons for successful implementation. These activities and functions are specified in Table 4.4 and will not be included in the analyses discussed in the remainder of the thesis as it focuses on the methods used for and results of routine implementation monitoring in Studio Schools. After the final pilot, PBL and coaching staff as well as students indicated they had no further concerns regarding completing the questionnaires.

4.4.3 Overview of original measures

Tables 4.3 and 4.5 contain the items that constitute the activities and functions that were monitored in the study during PBL and coaching sessions. Table 4.4 contains additional items that were asked once at the onset of a PBL project. Note that these measures contain the items and scales as these were hypothesised before these were used to collect any data. Based on collected implementation data, the psychometric properties of the measures were assessed upon which the measures were revised and finalised prior to data analysis. This process, as well as the definitive measures resulting from it, is reported in Chapter V.

Table 4.3: Overview of original scales and items used for measuring implementation of project-based learning lessons in Studio Schools

Scale	Items
<i>AA. Learning and developing skills as a means to an authentic end: Activities</i>	AA2. The problems or assignments central to project activities should be presented without structured guidance on where to start in order to complete it AA5. During the process of working on the project, students should work or meet with at least one adult with expertise that is of relevance to the project

<i>FA. Learning and developing skills as a means to an authentic end: Functions</i>	FA2. The problem central to the project should be of importance to students FA3. It should be clear to students and learning coach how studied subject content aligns with the end product(s) of the project FA4. Students' application of knowledge and skills reflects that of adult life FA5. The adults involved in the project should inspire students to put in their best effort
<i>AB. Engaging in a challenging enquiry process: Activities</i>	AB1. When assisting students, learning coaches should not present solutions to encountered problems AB2. When assisting students, learning coaches should not instruct students what they need to study to complete the project successfully AB3. Students should draw upon key methods of enquiry suitable with the project's discipline(s) (e.g., document analysis for English/History, collecting own data for Science)
<i>FB. Engaging in a challenging enquiry process: Functions</i>	FB1. Students should discover independently where/how the subject content and other factual information that is needed to understand/manage a problem and complete the project can be obtained FB2. Students should feel ownership of the project and its end product(s): it is their responsibility to come with the answers to complete it successfully
<i>AC. Obtaining and applying subject knowledge: Activities</i>	AC1. Students should draw on content of one or multiple subject(s) that are offered at Studio Schools (i.e. English language, English literature, math, science) whilst completing the project
<i>FC. Obtaining and applying subject knowledge: Functions</i>	FC1. It should be clear to students and teachers how content/knowledge that is needed to complete the project links to the subject content they require to learn to pass their exams
<i>AD. Explicit development of CREATE-skills: Activities</i>	AD1. During the process of working on the project, students should reflect explicitly on which of their CREATE skills they need to further develop to successfully complete the project
<i>FD. Explicit development of CREATE-skills: Functions</i>	FD1. It should be clear to a student and learning coach which CREATE skills a student needs to develop further to complete the project successfully FD2. It should be clear to a student and learning coach how students' project activities help them develop <u>specific</u> CREATE skills
<i>AE. Elaborating on subject content and reflecting on own work: Activities</i>	AE1a. Students should receive, as well as provide, explicit oral written feedback/critique on the products (deliverables) of their project activities from their classmates, including on the project's end product(s) AE1b. Students should receive, explicit oral written feedback/critique on the products (deliverables) of their project activities from their teachers, including on the project's end product(s) AE3. Classroom time should be spent as much as possible on working with subject content (independently developing/applying knowledge) and as little as possible on lecturing subject content

<i>FE. Elaborating on subject content and reflecting on own work: Functions</i>	FE1. Students should have sufficient opportunities to express thoughts and ideas about subject content to others (teacher/learning coach and fellow students) in the classroom and/or through assignments FE2. Students recognise those aspects of their own and others' work that can/should be improved FE3. Students recognise those aspects of their own and others' work that are particularly strong FE4. Students develop their own best way of learning/studying subject content for it to be of use towards project tasks and end products
<i>FF. Application/assessment of knowledge and skills in multiple ways and contexts: Functions</i>	FF1. It should be clear to students how knowledge of subject content attained in the context of the project could have utility in other contexts

Table 4.4: Overview of original scales and items used for measuring implementation of project-based learning structuring in Studio Schools

Scale	Items
<i>AA. Learning and developing skills as a means to an authentic end: Activities</i>	AA1. The project should have at least one concrete end product that will be exhibited publicly to an audience beyond the school AA3. One or more adults from outside the school with expertise/experience that is of relevance to the project should be involved in co-designing or commissioning the project AA4. One or more adults with expertise/experience that is of relevance to the project should be involved in providing instructional support and assessment towards project activities
<i>FA. Learning and developing skills as a means to an authentic end: Functions</i>	FA1. It should be clear to students and learning coach what the concrete product(s) of the project is/are
<i>FC. Obtaining and applying subject knowledge: Functions</i>	FC2. Over the scope of the project period, students should have covered an amount of National Curriculum similar to what they would have covered had they attended regular Academies
<i>AE. Elaborating on subject content and reflecting on own work: Activities</i>	AE2. Students should submit multiple drafts of the end product of the project and should explicitly reflect on each draft guided by feedback of the teacher/learning coach
<i>AF. Application/assessment of knowledge and skills in multiple ways and contexts: Activities</i>	AF1. During the process of working on the project, students should submit work and be assessed through a range of exercises.

Table 4.5: Overview of original scales and items used for measuring implementation of personal coaching sessions in Studio Schools

Scale	Items
<i>AA. Guided self-reflection on performance and behaviour: Activities</i>	AA1a. Student and coach should reflect on student's CREATE progress during coaching sessions AA1b. Student and coach should reflect on student's academic progress during coaching sessions AA2. Student and coach should reflect on student's behaviour

	AA3a. Student and coach should discuss the cues and situations that negatively affect student's performance at school/work AA3b. Student and coach should discuss the cues and situations that positively affect student's performance at school/work AA4a. During coaching sessions, a coach should guide the student to discovering methods for reinforcing constructive thought patterns and emotions AA4b. During coaching sessions, a coach should guide the student to discovering methods for reshaping dysfunctional thought patterns and emotions
<i>FA. Guided self-reflection on performance and behaviour: Functions</i>	FA1a. After a coaching session, students should be aware of their specific strengths with regards to their academic and CREATE progress FA1b. After a coaching session, students should be aware of their specific weaknesses with regards to their academic and CREATE progress FA1c. After a coaching session, coaches should be aware of students' specific strengths with regards to their academic and CREATE progress FA1d. After a coaching session, coaches should be aware of students' specific weaknesses with regards to their academic and CREATE progress FA2a. After a coaching session, students should be aware of their specific strengths with regards to their behaviour FA2b. After a coaching session, students should be aware of their specific weaknesses with regards to their behaviour FA2c. After a coaching session, coaches should be aware of students' specific strengths with regards to their behaviour FA2d. After a coaching session, coaches should be aware of students' specific weaknesses with regards to their behaviour FA3a. After a coaching session, students should be aware of the thought patterns that influence their perceptions of themselves and others positively FA3b. After a coaching session, students should be aware of the thought patterns that influence their perceptions of themselves and others negatively FA3c. After a coaching session, coaches should be aware of the thought patterns that influence students' perceptions of themselves and others positively FA3d. After a coaching session, students and coaches should be aware of the thought patterns that influence students' perceptions of themselves and others negatively. FA4a. After a coaching session, students should be aware of specific methods for reinforcing constructive thought patterns and emotions FA4b. After a coaching session, students should be aware of specific methods for reshaping dysfunctional thought patterns and emotions
<i>AB. Joint setting of goals/solutions and planning/reflecting on actions: Activities</i>	AB1. During coaching sessions, student and coach should discuss and determine concrete curricular and personal goals of the student. AB2. During coaching sessions, student and coach should discuss and determine the concrete actions that need to be undertaken to achieve these goals

	AB3. Student and coach should discuss the execution of planned actions to determine progress towards achieving the student's goals and (if necessary) to plan new actions
<i>FB. Joint setting of goals/solutions and planning/reflecting on actions: Functions</i>	FB1a. It should be clear to the student what the curricular and personal goals are that she/he is working towards FB1b. FB1. It should be clear to the personal coach what the curricular and personal goals are that the student is working towards FB2. It should be clear to students/personal coach what actions are to be undertaken by i) the student and ii) the coach to achieve these goals.
<i>AC. Self-direction towards goals and solutions: Activities</i>	AC1. During coaching sessions, coaches should use questions to help students self-examine their behaviours and attitudes AC2. During coaching sessions, coaches should not offer concrete advice or direct students towards solutions or actions.
<i>FC. Self-direction towards goals and solutions: Functions</i>	FC1a. After each coaching session, student/coach should feel that specified changes that are needed to improve student's behaviours/situation/learning were defined primarily by the student. FC1b. After each coaching session, student/coach should feel that specified changes that are needed to improve student's behaviours/situation/learning were not defined primarily by the coach.
<i>AD. Supportive relationships: Activities</i>	AD1. During coaching sessions, coach should provide sufficient opportunities for students to discuss personal matters (such as experiences at home, at school, and at work) without feeling exposed;
<i>AD. Supportive relationships: Functions</i>	FD1a. Students should trust their coaches FD1b. Coaches should trust their students FD2. Students and personal coaches should feel that the coach is aware of and understands students' personal characteristics FD3. Students should feel supported by their personal coach when engaging with new activities or (interactive) strategies with which they may be uncomfortable.

Final operationalisation

The measures presented in Tables 4.3 and 4.5 were operationalised into questionnaires to be rated during PBL and coaching sessions by the different rater groups. Table 4.6 presents the number of quantitative and qualitative items in each project-based learning questionnaire, and Table 4.7 presents the numbers for the personal coaching instruments. Quantitative questions could be rated in a 0-3 Likert-

scale, although some activities were rated on dichotomous (i.e. ‘Yes/No’) scales. As an example, Table 4.8 presents how PBL function FC1 was translated into a quantitative item for each rater group, and into a qualitative question for students and teachers. For a complete overview of the questionnaires used in the study, the reader is referred to Appendix B.

Table 4.6: Questionnaire composition by type of item per category of rater: project-based learning

Rater category → Type of item ↓	Researcher	Students	Teachers
N Quantitative items	20	24	21
Activities	10	10	10
Functions	10	14	11
N Qualitative questions	X*	3	3

**Only students and teachers completed qualitative questions*

Table 4.7: Questionnaire composition by type of item per category of rater: personal coaching

Rater category → Type of item ↓	Researcher	Students	Teachers
N Quantitative items	33	36	36
Activities	13	13	13
Functions	20	23	23
N Qualitative questions	X*	5	5

**Only students and teachers completed qualitative questions*

Table 4.8: Operationalisation example: PBL function FC1

Item	Researcher	Students	Teachers
Quantitative	<i>“It is clear to students/teacher how the content/ knowledge they study to complete their project links to that which they need to learn to pass their exams”</i> 0-3 Likert Scale	<i>“I don’t understand how the things I study to complete my project link to what I need to know for my exams”</i> 0-3 Likert Scale	<i>“It is not clear to me how the content/ knowledge that students study to complete their project links to the content/knowledge they need to learn to pass their exams”</i> 0-3 Likert Scale
Qualitative	<i>No qualitative questions answered by researcher</i>	<i>“Please describe what you have studied or learned today for this project and how you think it is important for what you need to know for your exams”</i> Free text answer	<i>“Please describe the subject content students have studied or learned today for this project, and how you think it aligns with the curriculum content that is required for students to obtain their qualifications”</i> Free text answer

4.5 Discussion

This chapter reported the process of developing measures for monitoring implementation of PBL and coaching in Studio Schools. An iterative process of micro-level programme theory development was followed by an operationalisation and piloting process in order to arrive at a set of instruments that were used to collect data for the study. In this section, this development process is further examined.

4.5.1 Programme theory development process

The process of specifying the micro-level programme theories for PBL and coaching built further on the inductive and user-focused approach to elucidating the Studio Schools programme theory as a whole that was reported in Chapter II. This implies that these theories are similarly driven by stakeholder, practitioner, and end-user input and less so by systematic assessment of the scientific literature. However, it needs to be noted here that core theoretical concepts discussed in the literature that were obtained through snowballing were included, as the activities and functions as well as the relationships amongst these described by participants were fortified or revised based on these concepts. This decision was made because common programme descriptions of project-based learning and personal coaching for secondary education purposes are lacking (Thomas, 2000), and in order to ensure that the specific practices and context of Studio Schools were captured as well as possible by the resulting measures. A more detailed reflection on the advantages and disadvantages of deductive (i.e. literature-driven), inductive, and user-focused approaches has been provided previously in Chapter II and will not be discussed further here. The explicit process of mapping activities to functions resulted in very detailed theories of change that assume rather linear relationships between different activities and functions. In reality, as was argued in paragraph 4.3.2, it is unlikely that this relationship is this

linear in nature (Bonell et al., 2012), and it was expected that the programme theories would be revised as a result of using the measures in the schools during the study (as demonstrated in the next chapter). Nevertheless, it is argued here that the detailed specification of programme theories at the micro level –including the explicit matching of activities and functions– is beneficial to programme evaluation regardless of the accuracy of the theory as originally specified. Only through explicit and detailed evaluation of a programme theory can it be verified or falsified by empirical research (Michie & Prestwich, 2010; Moore et al., 2014). However, it is acknowledged that the more detailed the specification of a theory, the greater the likelihood that it will need some revision once evaluated, as it will contain more assumptions that can be explicitly evaluated. This may require a substantial amount of time and other important resources.

4.5.2 Operationalisation process and types of data collected

A major strength of this study results from the extensive iterative process of development and piloting of the used measures. Measures were created and revised based on stakeholder input, as well as feedback received from experts. These were then piloted in two participating Studio Schools and revised based on this process as well as feedback obtained from participants. However, the operationalisation of implementation fidelity in this study is also subject to some limitations. First, it was decided to measure implementation exclusively quantitatively. It is recognised that incorporating qualitative data in process evaluations offer substantial benefits to the utility of these. In particular, the combination of qualitative and quantitative methods can help increase understanding of outcomes and improve interventions and their implementation (Moore et al., 2014). For example, if quantitative results indicate that fidelity of implementation of specific activities or functions is poor, interviews and or

focus groups with providers and participants may indicate barriers to implementation or uptake and increase understanding of why activities may not always trigger functions for each participant. As a result, qualitative data are commonly included in process evaluations to increase their explanatory power (Audrey et al., 2006; Hawe et al., 2004b; Hargreaves et al., 2010; Moore et al., 2014). Qualitative data –other than the qualitative responses provided by teachers and students for the purpose of rating by the researcher– were not collected as part of the main implementation study reported in this thesis. The most important reason for this is the limited amount of resources available for the study. Scheduling and conducting (focus) group interviews would require additional visits to the schools, which proved very difficult to arrange (see Chapters V and VI). Collecting qualitative implementation data through teacher reports (e.g., logs or diaries) was considered, but found too much of a burden to staff in discussions with school principals. An additional reason for not including qualitative data in the study was its demonstrative nature. The main concern of the project is to assess the feasibility of a function-focused approach towards standardising complex interventions and monitoring their implementations. Although qualitative data play a very important role in process evaluation, fidelity monitoring is essentially a quantitative exercise. Therefore, in order to concentrate efforts and resources on core research questions, a focus was given to collecting and analyzing quantitative data. An additional potential limitation of the study is posed by the operationalisation of quantitative form- and function-focused fidelity measures in questionnaires and checklists. Although such measures form the most commonly used in fidelity monitoring (Breitenstein et al., 2010; Durlak & DuPre, 2008), it is possible that some of the studied constructs could have been more appropriately measured through other forms of measurement. For example, some functions of PBL and

coaching include attitudes and motivation structures of students. By asking students to report on these explicitly in questionnaires directly after implementation of a PBL or coaching session, there is a chance that their answers do not accurately represent their true attitudes and/or motivations. Qualitative interviews in which students would be asked richer questions about their experiences at their schools could have been one alternative method of obtaining data related to these constructs that could still be coded quantitatively. Although triangulation of data sources (i.e., the researcher, students, and staff) aimed to provide some assurance regarding the validity of responses, the possibility that using alternative methods for collecting the data would result in different findings cannot be ruled out entirely. Considering the most appropriate instruments for collecting form- and function-focused fidelity data is an important consideration for any future research conducted on the topic or utilising these concepts in process evaluations.

4.6 Chapter summary

In this chapter, the process of developing implementation measures based on the concepts defined in Chapter III was discussed. It explained the focus of this thesis on two school-based elements of the Studio School model (project-based learning and personal coaching only). A basic theoretical framework was next developed for both elements based on three sources: a non-systematic ('snowball') review of the scientific literature, notes made during presentations given by the Studio Schools Trust (SST) about PBL and coaching, and a one-hour focus group with members of the SST. Core components were distilled for both elements and subsequently mapped to the Studio Schools logic model that was developed in Chapter II. Micro-level logic models were created for both elements and were revised based on feedback received

by the SST. Once the micro-level logic models were approved, activities and functions were specified for PBL and coaching based on the three information sources described above and were integrated into the micro-level logic models. All PBL and coaching activities and functions were operationalised as quantitative questionnaire items, with additional qualitative questions asked from students and teachers of which the answers were coded by the researcher. The purpose of this exercise was gain data on some functions (e.g., thought patterns) that could not be observed in vivo by the researcher. The chapter finally presented the measures used for monitoring the implementation of PBL and coaching as these emerged from their development process, but prior to revision upon an assessment of their psychometric properties. This exercise will be reported in Chapter V, alongside a description of the process of data collection for the remainder of the project.

Chapter V: Assessing the Validity and Reliability of Used Measures

5.1 Introduction

In the previous chapter, the process of specification and subsequent operationalisation of implementation measures for Studio Schools was described. These measures, which were used for the monitoring of project-based learning (PBL) and personal coaching in the schools, focus on the activities as well as the functions of these elements. The current chapter is concerned with an assessment of the reliability and validity of these measures.

5.1.1 Research question

This chapter addresses the second research question of the DPhil thesis:

Can implementation of complex social interventions be measured validly and reliably in terms of the theoretical functions (next to the form) of their components?

5.2. Data collection

All data on the implementation of PBL and coaching were collected in Studio Schools between September 2014 and January 2015. The limited timeframe set for the completion of a doctoral programme prevented the collection of separate data for testing the validity and reliability of the used measures and subsequent testing of fidelity. As a result, the data that were used to complete this assessment are the same that were used to assess the fidelity with which PBL and coaching are currently implemented in the participating Studio Schools. It is noted here that this approach carries some limitations (DeVellis, 2003), which will be further discussed.

5.2.1 Settings and sampling

Schools

Four Studio Schools geographically located across the South and Midlands of England (counties are not included here to preserve the anonymity of the schools) were selected to participate in the study. Convenience sampling was used to select participating schools. The Studio Schools Trust nominated candidate schools in consultation with FvU based on the following criteria: willingness to participate in the study, and general representativeness of the settings and geographic distribution of Studio Schools. In addition, the sample needed to consist of schools that newly opened in September 2013 as well as schools that had been open for one or more years at that time. The small number of participating schools and the nature of the sampling technique were considered the maximum limit for feasibility, but cause for some limitations to the study that are further discussed in section 5.6 of this chapter.

Sessions

Project-based learning

A total of twelve PBL sessions (three in school 1, four in school 2, three in school 3, and two in school 4) were sampled for data collection, and PBL implementation data were collected during eleven visits to the schools. Dates for data collection were sampled on a basis of convenience and agreed with the principals of the schools based on suitability with the schedules of the researcher and the schools. Sessions were sampled per school and also by means of convenience sampling in consultation with the principal and the teachers responsible for any PBL sessions taking place on the dates at which the visits took place. Potentially suitable sessions were identified by the researcher based on the following criteria: participating students were in Year Ten

or Year Twelve at the time of their participation⁵, and –where possible– the session was not facilitated by a teacher who had already facilitated a previously monitored session. The researcher then proposed sessions to the principal and responsible teachers, who declined in a small minority of cases. Reasons for declination included an excessive burden to the facilitating teachers or the school as a result of other commitments, or a lack of actual project-based learning pedagogy used during the session as a result of a revised lesson plan.

All project-based learning sessions took place within the respective Studio Schools, and started in a single classroom. Students were typically given a plenary introduction to their assignments at the onset of a new project, or instructions regarding where to continue in case a previously initiated project continued during the session. After receiving an introduction, students worked on their projects independently or in groups and would sometimes break out to work on their projects elsewhere in the schools.

Personal coaching

A total of 36 personal coaching sessions (twelve in school 1, fourteen in school 2, six in school three, and four in school 4) were sampled for data collection. Coaching data were collected during eleven visits to the schools. In most cases, visits set up for the purpose of collecting PBL data were also used to collect coaching data. In each school, sessions were sampled by means of convenience sampling in consultation with the personal coaches that were coaching on the dates at which the visits took place. Coaches were requested to adhere to their planned coaching schedule as much

⁵ Only students enrolled in Year Ten and Year Twelve were included in the monitoring of project-based learning, because students in these year groups newly joined their Studio School and had no previous experience with this pedagogy.

as possible, but were allowed to opt out students that they considered too vulnerable to be eligible for participation. Selected sessions were included for data collection if consent was obtained from the participating student and coach.

All personal coaching sessions took place within the respective Studio Schools. In the majority of cases, sessions were held in designated private rooms. In a small minority of cases, sessions were held in an open area but without the presence of other students or staff members.

Participants

Ethics

At the onset of the study (i.e. when students' background data were collected) all participants in the study were provided with written information sheets explaining the voluntary nature of their participation, their right to not participate or withdraw from participation at any time, and the absence of any consequences related to their decision to participate or not participate in the study. The doctoral candidate repeated this information verbally at the start of plenary meetings in which the background data of participating students were collected. Written consent was then obtained from all participating students and teachers. Students were provided with letters for their parents, in which the same information was provided that was given to the students and in which parents were informed of their right to opt out or withdraw their child from the study at any time. At the beginning of any monitored project-based learning session, teachers and students were reminded that their participation was voluntary and that they could stop participating at any time. Because personal coaching sessions were recorded, explicit written consent for any recording was obtained from participating coaches and students on top of their general consent to participate in the

study. Participants were reminded of their right to stop a recording at any time or to request deleting of the recording after the session was completed. Ethical approval for the study was obtained from the University of Oxford Social Sciences & Humanities Inter-Divisional Research Ethics Committee (Ref No: SSD/CUREC2/13-18).

Sampling strategy

- Project-based learning

Once a project-based learning session was selected for data collection, all students and teachers who participated in the session were invited to participate in the study, provided they had given their written informed consent at the onset of the study. At the beginning of each PBL session, all participants were given a plenary explanation of the voluntary and confidential nature of their participation and their right to not participate. All participants from whom written consent had been obtained and who completed their questionnaire were included in the study. In most of the sessions, the majority of students participated and completed their questionnaires.

- Personal coaching

Because personal coaching in Studio Schools is a one-to-one activity, participants in selected coaching sessions were automatically recruited in parallel with the session they participated in. This implies that if a personal coaching session was selected once consent from student and coach was obtained, both participants were included.

Overview of participants

- Project-based learning

Table 5.1 presents an overview of the number of sessions, students, and teachers that were included for data collection on project-based learning in the participating schools. Note that separate values are included for the number of observations (i.e.

completed questionnaires) that were obtained from students and teachers, and the number of the unique students and teachers that provided these data. Details on participants' gender and (in the case of students) year groups are provided in Tables 5.2. and 5.3. As can be seen, female students and teachers are slightly over-represented in the sample, although the latter figure is in line with the overrepresentation of women in secondary school teaching in England (ONS, 2014a). Girls were also slightly overrepresented in the sample of students who completed the PBL questionnaires when compared to the overall sample of students included in the study at the onset, although the overrepresentation did not exceed nine percent (see Table 6.6 in Chapter VI for the overall included sample). Because the only data available are those collected as part of the study, it is impossible to compare these characteristics with those of the full student bodies of the participating schools (i.e. those not participating in the study) or those of other Studio Schools. Participants' further demographic characteristics will be presented in detail in Chapter VI.

Table 5.1: Overview of sessions, students, and teachers included per school in PBL data collection

School	1	2	3	4	Total
PBL sessions (N)	3	4	3	2	12
Students – observations (N)	66	24	25	30	145
Students – unique (N)	45	18	25	30	118
Teachers – observations (N)	3	4	3	2	12
Teachers – unique (N)	2	2	3	2	9

Table 5.2: Overview of gender and year groups of students included per school in PBL data collection

School	1	2	3	4	Total
Girls (%)	44.4	50.0	80.0	61.3	57.1
Year group (N)					
10 (N)	45	18	13	7	83
12 (N)	0	0	12	23	35

Table 5.3: Overview of gender of teachers included per school in PBL data collection

School	1	2	3	4	Total
Female (% observations)	33.3	50.0	66.7	100	58.3
Female (% unique)	50.0	50.0	66.7	100	66.7

- Personal coaching

An overview of the personal coaching sessions, students, and coaches that were included per school for data collection on personal coaching is provided in Table 5.4. As for Table 5.1, separate values are provided for the number of observations obtained from participants and the number of unique participants that were included. Note that the number of students presented in Table 5.4 does not match the number of coaches, as one student did not provide consent for completing the questionnaire (the student did however provide consent for a recording and for completion of the questionnaire by the coach). Tables 5.5 and 5.6 present details on participants' gender and year groups. Note that, contrary to the decision made for data collection on project-based learning, students in Year Eleven and Year Thirteen were included alongside students in Year Ten and Year Twelve for data collection on personal coaching, because it proved unfeasible to recruit enough students for the study while including only Year Ten and Year Twelve students.

Table 5.4: Overview of sessions, students, and coaches included per school in coaching data collection

School	1	2	3	4	Total
Coaching sessions (N)	12	14	6	4	36
Students – observations (N)	12	13*	6	4	35
Students – unique (N)	12	13	6	4	35
Teachers – observations (N)	12	14	6	4	36
Teachers – unique (N)	3	5	4	2	14

*One student did not complete the questionnaire

Table 5.5: Overview of gender and year groups of students included per school in coaching data collection

School	1	2	3	4	Total
Female (%)	50.0	35.7	83.3	50.0	51.4
Year group (N)					
10	4	3	1	0	8
11	8	10	0	0	18
12	0	0	2	2	4
13	0	0	3	2	5

Table 5.6: Overview of gender of coaches included per school in coaching data collection

School	1	2	3	4	Total
Female (% observations)	100	71.4	83.3	100	86.1
Female (% unique)	100	60.0	75.0	100	78.6

Differences between the numbers of included sessions per school

It was originally planned to sample an equal number of project-based learning and personal coaching sessions per school. It was also intended to sample a much larger number of sessions across all schools. However, several factors interfered with this plan, resulting in discrepancy between the numbers of sessions that were included in the study per school. These factors and their implications in terms of limitations to the study are discussed in paragraph 5.6.4 of this chapter.

5.2.2 Procedure

Procedure for collecting data on project-based learning

At the beginning of each PBL session, the facilitating teachers briefly introduced the researcher to participating students, after which the session began. During the sessions, the researcher completed a section of the 20-item checklist presented in Appendix B1 based on in-vivo observation of the session. The researcher coded qualitative responses received by the teachers based on the remainder of his checklist after completion of a data collection visit. Once the session was completed, all students that had provided their consent to participate in the study were invited to complete the 21-item checklist presented in Appendix B2. All students who completed the questionnaire did so in the classroom in which the session had initiated, before they left for their next lesson. Teachers were invited to complete the 21-item checklist presented in Appendix B3, and typically completed the questionnaire while students were completing theirs. In some cases, due to a lack of time between a

project session and the next lesson, teachers completed their questionnaire at a later point during the data collection visit.

Procedure for collecting data on personal coaching

Prior to the start of each personal coaching session, participating students and coaches were explained the nature and purpose of their involvement, its voluntary character, and their right to stop participating at any time. Once consent had been obtained from both participants, the researcher installed a small video recording device (GoPro HERO 3+) and an audio recording device (Olympus VN-712PC) in the coaching room, after which the session began. At the end of the session, students and coaches were invited to complete the 37-item checklists presented in Appendices B5 and B6 respectively. Participants typically completed their questionnaire immediately after a coaching session had ended. However, in some instances, they completed their questionnaires at a later point during the data collection visit. All coaching sessions were rated retrospectively by the researcher using the 35-item checklist presented in Appendix B4, based on recordings made of the sessions with the equipment described above.

5.3 Construct validity and internal consistency

5.3.1 Methods and measures

All implementation data were collected using paper-based quantitative questionnaires in which participants (teaching staff and students) and the researcher rated items on ordinal scales. Therefore, measures described below and in section 5.3 were selected based on their appropriateness for demonstrating the validity and reliability of ordinal-level measures.

Methods and measures used for determining construct validity

Validity refers to the extent to which (aspects of) research provide an accurate representation of the real world. One aim of this study is to assess the extent to which functions can be measured with validity next to form as part of implementation monitoring. The core dimension of validity relevant to this exercise is that of *construct validity*. Construct validity refers to the validity of a measure of some attribute or quality that is not operationally defined (Cronbach & Meehl, 1955) and that needs to be constructed from other operationally defined items. In order to determine if the factor structure hypothesised in Chapter 4 could be validated empirically, including the hypothesis that activity items and function items would load on separate factors, exploratory and confirmatory factor analyses were conducted. Factor analysis was selected as the procedure of choice because it allows an analysis of the latent constructs (i.e. activities and functions) that account for the variance in measured items. Thus, it allows for an assessment of the extent to which measurable items load together onto aggregated activities and functions that represent latent programme components, which is key to determining the feasibility and potential validity of this approach. Given the exploratory nature of the doctoral project, factor analyses were only conducted on datasets that allowed for these in terms of their sample size. Therefore, factor analyses (exploratory and confirmatory) were only used to analyse PBL implementation data provided by students (N= 145), whilst analytic techniques requiring less statistical power were used for the other datasets (see Table 5.7). All factor analyses were conducted in R version 3.1.3 for Mac: packages used were polycor V0.7-8 and psych V1.5.1 for all exploratory factor analyses and lavaan V0.5-18 for confirmatory factor analysis.

Exploratory factor analysis

The literature recommends multiple methods for conducting exploratory factor analyses and selecting appropriate factor structures based on their results (e.g., Courtney, 2013; Williams et al., 2012). For this project, a three-step approach was chosen. First, polychoric correlation coefficient tables (appropriate for assessing the correlation between ordinal-level indicators) were generated to inform an initial assessment of the correlation between items included in the student PBL implementation questionnaires. Then, after revisions (i.e. the removal or recoding of items, see Figure 5.1 for a flowchart of procedures and Table 5.10 for an overview of recoded and removed items) were made, and a Scree plot plotting eigenvalues was generated (see Figure 5.2) to determine a suitable number of factors. Finally, orthogonal varimax rotation was used to determine empirically supported factor structures for the PBL instruments. Orthogonal rotation operates under the assumption that factors are uncorrelated, versus oblique rotation, which assumes correlation. The former was selected to enter the analysis from a neutral point of view: assuming factors were uncorrelated unless proven otherwise.

Confirmatory factor analysis

Confirmatory factor analysis was conducted through the specification of a measurement structural equation model (SEM). Recommendations for appropriate sample sizes to support SEM vary, but an absolute sample size of over 100 cases and a case-to parameter ration of 10:1 are suggested a minimum for a valid analysis (Bowen & Guo, 2011). As a result, the only data set collected within this project that could possibly provide an adequate sample size for a confirmatory factory analysis is the set provided by students' ratings of the implementation of project-based learning (see Table 5.7). The size of this sample (N=145) is still below desirability (due to a

suboptimal case-to parameter ration of below 10:1: see Kline, 2011; Williams et al., 2012), causing some limitations that are discussed in paragraph 5.6.5. In addition, the small sample size does not allow for accounting for clustering (of responses within sessions, within schools), of which the implications will be discussed later in this chapter. The factor structure that was suggested appropriate for monitoring PBL in Studio Schools by the previous exploratory factor analyses was specified as a structural equation model and estimated using weighted least squares estimation as recommended for ordinal and other non-normal data (Bowen & Guo, 2011). To compare the fit of the suggested factor structure to that of alternative factor structures (e.g., structures in which activities and functions load together on the same factors), alternative models were specified and evaluated according to the same procedure. Measures of fit were selected to provide a complete picture of the model's fit. The Chi-square (X^2) p-value was used to test the exact-fit hypothesis (that is: the model does or does not represent an accurate representation of the real factor structure, implied by p-value higher or lower than 0.05 respectively). The Root Mean Square Error of Approximation (RMSEA) was used to demonstrate an approximate goodness of fit, with values closer to 0 representing better fit. The RMSEA 90% confidence interval was also calculated. The Comparative Fit Index and Tucker-Lewis Index (TLI) were used to estimate the explanatory power of specified models versus a default situation in which no model is specified, with values closer to 1 representing better fit.

Table 5.7: Overview of analysis procedures used to determine the construct validity of used measures

Component of focus	Analysis	Data analysed
Project-based learning	<u>Exploratory Factor Analysis</u>	
	<i>Assessment of polychoric correlations</i>	Student implementation questionnaires
	<i>Scree plot inspection</i>	Student implementation questionnaires
	<i>Orthogonal varimax rotation</i>	Student implementation questionnaires
	<u>Confirmatory Factor Analysis</u>	
	<i>Specification of SEM with WLS estimation</i>	Student implementation questionnaires
	<i>Comparison of model fit for alternative hypothesised factor structures</i>	Student implementation questionnaires
Personal coaching	<i>Assessment of polychoric correlations</i>	- Student implementation questionnaires - Teacher implementation questionnaires - Researcher implementation questionnaires

Methods and measures used for determining internal consistency

Apart from assessing the validity (specifically the construct validity) of the used measures, this chapter is concerned with their reliability. The reliability of a measure refers to the extent to which it can be expected to produce similar results under the same conditions when used repeatedly. One dimension of measure reliability is its internal consistency. The internal consistency reliability of a measure is the extent to which items that form part of the same scale (i.e. measure the same construct) produce similar results, and can thus be treated empirically as the same scale. To determine the internal consistency of the scales and subscales used to monitor the implementation of PBL and coaching in Studio Schools, Cronbach's alphas (α) were calculated for all scales and subscales used in all instruments (see Table 5.8). Internal consistency was determined prior to a revision of the factor structure and the scales as well as after revision to assess if the revision had a desirable influence. All internal consistency analyses were carried out in SPSS version 22 for Mac.

Table 5.8: Overview of analysis procedures used to determine the internal consistency of used measures

Component of focus	Analysis	Data analysed
Project-based learning	<i>Initial assessment of internal consistency (Assessment of Cronbach's α) (Table 5.9)</i>	- Student implementation questionnaires - Teacher implementation questionnaires - Researcher implementation questionnaires
	<i>Reassessment of internal consistency (After revision of instruments) (Table 5.16)</i>	- Student implementation questionnaires - Teacher implementation questionnaires - Researcher implementation questionnaires
Personal coaching	<i>Initial assessment of internal consistency (Table 5.17)</i>	- Student implementation questionnaires - Teacher implementation questionnaires - Researcher implementation questionnaires
	<i>Reassessment of internal consistency (After revision of instruments) (Table 5.19)</i>	- Student implementation questionnaires - Coach implementation questionnaires - Researcher implementation questionnaires

5.3.2 Procedure

The construct validity and internal consistency of the used implementation measures were evaluated in four steps, using the methods and measures described in the previous paragraph (see Figure 5.1). For project-based learning, a fifth step was added in the form of a confirmatory factory analysis of the student-reported PBL data. First, an initial assessment of the internal consistency was completed for all scales and subscales comprising the instruments that were used by the three rater groups to rate the implementation of PBL and coaching. Second, exploratory factor analyses were carried out after which some items were removed from the PBL and coaching checklists because of clear misfit (see Table 5.10). This is in line with the recommendation that mismatching items be removed from scales or recoded depending on the size and their direction with other items upon psychometric

assessment (DeVellis, 2003). It is noted here and further discussed in paragraph 5.6.3 that these modifications de facto represent a refinement based on empirical findings of the original programme theories of these components. An additional set of EFAs formed the third step of the process and led to an empirically supported factor structure for both measures. Finally, the internal consistency of all scales was reassessed under the revised factor structure. The validity of the factor structure of the student-reported measure of PBL implementation was further assessed using the confirmatory factor analysis procedure described in the previous paragraph.

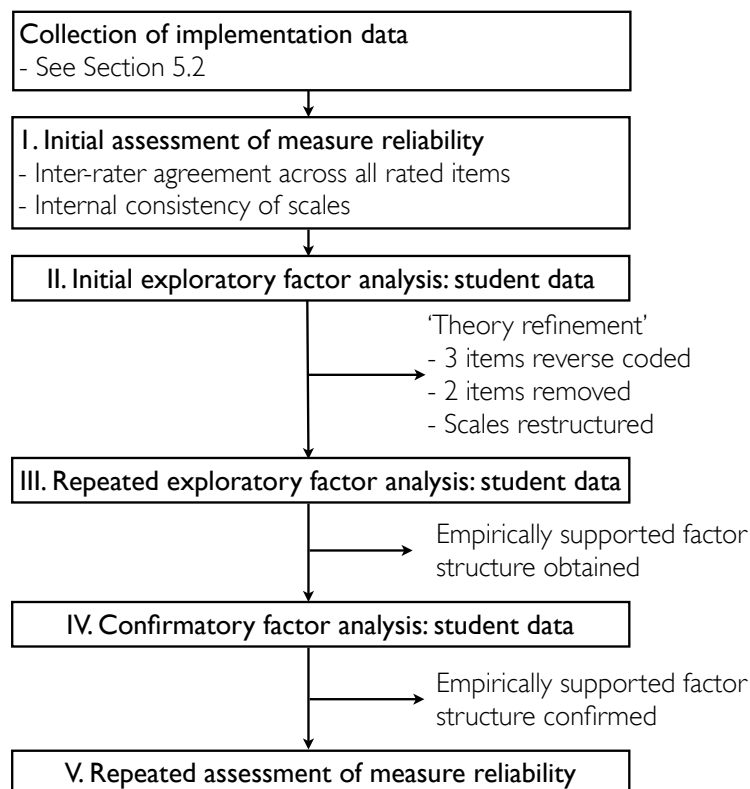


Figure 5.1: Flowchart depicting analysis procedure for determining the validity and internal consistency of used measures

5.3.3 Findings

Project-based learning

Initial assessment of internal consistency

The first step in determining the validity and reliability of the measures used for monitoring the implementation of PBL in Studio Schools consisted of assessing the

internal consistency of the original PBL measures as described in paragraph 4.4.3 (see Appendix B for an overview of the complete questionnaires). The results of this analysis are presented in Table 5.9.

Table 5.9: Internal consistency statistics for original scales and subscales used to monitor project-based learning in Studio Schools

Scale	α (N items) - Teacher	α (N items) - Students	α (N items) - Researcher
Activities A	.000 (2)	-.190 (2)	.136 (2)
Functions A	.283 (3)	.498 (4)	.148 (4)
Activities B	-.490 (3)	-.347 (3)	-1.182 (3)
Functions B	.710 (2)	.541 (2)	X* (1)
Activities C	X* (1)	X* (1)	X* (1)
Functions C	X* (1)	X* (1)	.231 (2)
Activities D	X* (1)	X* (1)	X* (1)
Functions D	.573 (2)	.686 (2)	X* (0)
Activities E	.871 (3)	.581 (3)	.495 (3)
Functions E	.553 (2)	.532 (4)	X* (1)
Functions F	X* (1)	X* (1)	.570 (2)
All Activities	.621 (10)	-.056 (10)	.443 (10)
All functions	.843 (11)	.806 (14)	.605 (10)
Overall	.818 (21)	.747 (24)	.750 (20)

* No α 's were calculated for scales comprising a single item

As becomes clear from Table 5.9, the presented α 's for some activity subscales as well as the overall activity scales used for the student- and researcher-reported PBL measures were suboptimal. Although the overall function scales appear to be much more consistent, this cannot be said for some of the function subscales. These results indicate that, for all rater groups, some items were inconsistent with other items under the factor and scale structure that was originally hypothesised (see Tables 4.3 and 4.5 in paragraph 4.4.3. for the original PBL and coaching factor structures).

Exploratory factor analysis

An initial assessment of internal consistency suggested a mismatch between some items, most notably in the activities scale of component B ('Engaging in a challenging enquiry process'). The first exploratory factor analysis on student reported data confirmed this problem. Furthermore, it brought to the light an additional situation

that required attention: some items appeared not to load at a significant level with any other item. The polychoric correlations table that was generated as part of the first exploratory factor analysis is included in Appendix C1. Key decisions regarding the recoding and removal of items that were made following inspection of the polychoric correlations between items are presented in Table 5.10. Note that these decisions were made to arrive at a measure that is as internally consistent and valid as possible, but represent a diversion from the theory on which the original measure was based. This has some implications for analysis, which are presented later in this chapter.

Table 5.10: Recoding/removal of PBL items poorly correlated with overall structure in student questionnaire

Original item name	Action	Reason
Activity A2 (AA2)	Reverse coded	Strongly* negatively related to other items (highest correlation = -0.52)
Activity A5 (AA5)	Removed	Not substantially related to any other item (highest correlation = 0.30)
Activity B1 (AB1)	Reverse coded	Strongly negatively related to other items (highest correlation = -0.60)
Activity B2 (AB2)	Reverse coded	Strongly negatively related to other items (highest correlation = -0.50)
Function E1 (FE1)	Removed	Not substantially related to any other item (highest correlation = 0.27)

* Coefficients were interpreted as recommended by Hair et al (1995): 0.30 = minimal, 0.40 = important, 0.50 = practically significant

After the revisions described in Table 5.10 were made, a new polychoric correlations table was generated based on the revised item set (see Appendix C2). Inspection of the new table suggested that the factorability of the polychoric correlation matrix was sufficient to proceed (i.e. enough correlations above .30 between included items were present: see Hair et al., 1995; Williams et al., 2012). A Scree plot was generated (see Figure 5.2), which suggested an optimal factor structure of three factors. Orthogonal varimax rotation confirmed the suitability of this structure (see Table 5.11), but also suggested a four-factor structure (see Table 5.12) that explained a slightly larger cumulative amount of variance. It further suggested that items representing activities

and functions loaded mostly on a single factor with items of their respective categories, although in some cases items loaded well onto multiple factors. The factor structures presented in Tables 5.11 and 5.12 were compared in confirmatory factor analyses.

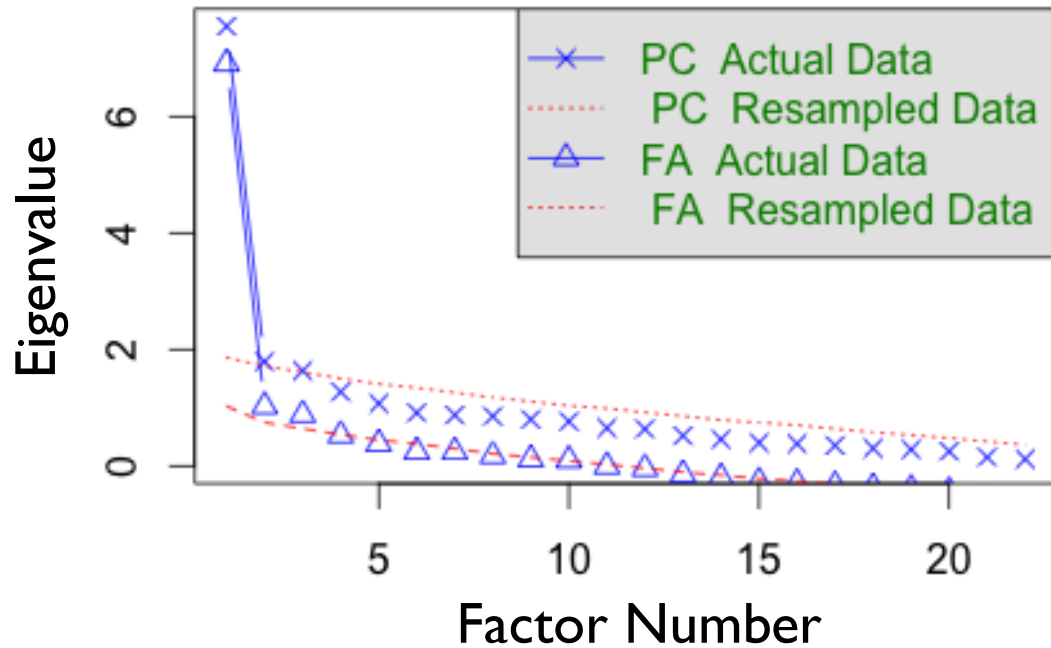


Figure 5.2: Scree plot suggesting a factor structure with three factors for the student PBL questionnaire. FA= Factor Analysis, PC= Principal Components. R resamples data automatically for the purpose of parallel analysis, a recommended procedure for factor extraction (Baglin, 2014).

Table 5.11: Factor loadings after orthogonal varimax rotation for three factors for the student PBL questionnaire

Original item name*	Factor 1	Factor 2	Factor 3
Activity A2 (AA2)	0.368	0.391	0.307
Activity B1 (AB1)	0.310	0.668	0.144
Activity B2 (AB2)	0.193	0.661	
Activity B3 (AB3)	0.215	0.558	0.125
Activity C1 (AC1)	0.232	0.598	
Activity D1 (AD1)	0.494	0.135	-0.122
Activity E1a (AE1a)	0.213	0.575	
Activity E1b (AE1b)	0.156	0.683	0.194
Activity E3 (AE3)	0.228	0.521	
Function A2 (FA2)	0.511	0.450	0.207
Function A3 (FA3)	0.479	0.187	0.261
Function A4 (FA4)		0.164	0.519
Function A5 (FA5)	0.512	0.261	0.134
Function B1 (FB1)	0.384	0.281	0.159
Function B2 (FB2)	0.561	0.256	0.363
Function C1 (FC1)			0.507

Table 5.11: Factor loadings after orthogonal varimax rotation for three factors for the student PBL questionnaire – continued from previous page

Original item name*	Factor 1	Factor 2	Factor 3
Function D1 (FD1)	0.621	0.324	-0.336
Function D2 (FD2)	0.756	0.274	
Function E2 (FE2)	0.526	0.203	0.280
Function E3 (FE3)	0.641	0.233	0.223
Function E4 (FE4)	0.473	0.381	0.185
Function F1 (FF1)	0.300		0.630

*AA1, AA3, AA4, AE2, FA1, FC2, and AF1 were not measured using the student PBL questionnaire

	Factor 1	Factor 2	Factor 3
Eigenvalue	3.916	3.696	1.687
Variance (proportion)	0.178	0.168	0.077
Variance (cumulative)	0.178	0.346	0.423

Table 5.12: Factor loadings after orthogonal varimax rotation for four factors for the student PBL questionnaire

Original item name*	Factor 1	Factor 2	Factor 3	Factor 4
Activity A2 (AA2)	0.375	0.469		0.131
Activity B1 (AB1)	0.665	0.325	0.149	
Activity B2 (AB2)	0.634	0.152	0.181	
Activity B3 (AB3)	0.551	0.201	0.143	
Activity C1 (AC1)	0.593	0.254		
Activity D1 (AD1)	0.105	0.155	0.564	
Activity E1a (AE1a)	0.564	0.144	0.185	
Activity E1b (AE1b)	0.686	0.272		
Activity E3 (AE3)	0.513		0.344	0.230
Function A2 (FA2)	0.438	0.506	0.237	
Function A3 (FA3)	0.181	0.578	0.106	
Function A4 (FA4)	0.161			0.863
Function A5 (FA5)	0.253	0.447	0.284	
Function B1 (FB1)	0.275	0.333	0.216	0.125
Function B2 (FB2)	0.250	0.544	0.272	0.267
Function C1 (FC1)		0.195		0.535
Function D1 (FD1)	0.300	0.192	0.695	-0.199
Function D2 (FD2)	0.233	0.509	0.617	0.110
Function E2 (FE2)	0.202	0.639	0.113	
Function E3 (FE3)	0.209	0.608	0.319	
Function E4 (FE4)	0.368	0.520	0.170	
Function F1 (FF1)		0.550		0.408

*AA1, AA3, AA4, AE2, FA1, FC2, and AF1 were not measured using the student PBL questionnaire

	Factor 1	Factor 2	Factor 3	Factor 4
Eigenvalue	3.521	3.456	1.847	1.446
Variance (proportion)	0.160	0.157	0.084	0.066
Variance (cumulative)	0.160	0.317	0.401	0.467

Confirmatory factor analysis of student PBL data

Multiple models were specified as measurement structural equation models to determine which model provided the best fit to the student-reported data (N= 145, of which 114 were included for analysis). Compared models were designed to confirm or falsify theoretical assumptions and included structures with various numbers of factors. These were informed by exploratory factor analysis (which suggested that three- as well as four-factor structures could fit the data) and the PBL programme theory and micro-level logic model developed in Chapter IV (which informed the specification of the factor structures: see Table 4.1 and Figure 4.3), and included structures in which activities and functions were specified to load onto separate factors as well as structures in which they were allowed to load onto the same factors. Two of these models –here referred to as Model 1 and Model 2– ‘passed’ the thresholds of the relevant fit statistics, indicating these provide an appropriate fit with the data. No models passed the relevant fit tests when accounting for clustering (of responses within sessions and within schools) which was anticipated considering the small sample size. Failure to account for clustering in CFA may compromise the validity of a demonstrated model fit as well as the values of its parameter estimates if clustering among responses is high (Pornprasertmanit et al., 2014). However, an item-by-item analysis of the intra-class correlation coefficients (ICCs) of the data reported by students revealed low very low levels of clustering at the school level (average= .02, range= <.0001 - .12) and the session level (average= .03, range= <.0001 - .15). A scale-by-scale (i.e. clustering at the factor-level) showed a similar image (school-level average= .02, range = <.0001 - .05; session-level average= .06 , range = <.0001 - .12). Hence it was decided to carry out the analyses at the individual level for their demonstrative value. Both models included a four factor structure with one factor

comprising mostly (Model 1) or exclusively (Model 2) activities, and the remaining three factors comprising mostly (Model 1) or exclusively functions (Model 2). The three-factor model suggested by Figure 5.2 did not provide an appropriate fit with the data. The reader is referred to Appendix C4 for a full overview of all models that were compared. Table 5.13 provides the fit statistics generated for both models that passed the fit tests. Tables 5.14 and 5.15 provide an overview of the factor structures and parameter estimates for both models.

Table 5.13: Fit statistics for models that provided significant fit with student-reported PBL data (N=114/145), using WLS estimation*

Statistic	Model 1	Model 2
X ² P-value	0.113	0.352
Degrees of freedom	203	164
CFI	0.995	0.999
TLI	0.994	0.998
RMSEA (90%** CI)	0.032 (0.000-0.052)	0.018 (0.000 – 0.045)
RMSEA P-value < 0.05	0.933	0.979

*WLS estimation excludes cases with missing data from analysis

**A confidence interval of 90% is conventional for SEMs (Kline, 2011)

Table 5.14: Parameter estimates for Model 1

Parameter	Factor Loadings (Unstandardised)	SE	Factor Loadings (Standardised)	ME Variances*	Thresholds
<i>Factor 1</i>					
AB1	1.000		0.796	0.367	-1.845 / -0.951 / 0.237
AB2	0.884	0.083	0.703	0.505	-1.657 / -0.643 / 0.386
AB3	0.791	0.091	0.629	0.604	-1.581 / -0.453 / 0.694
AC1	0.841	0.082	0.669	0.552	-1.743 / -0.431 / 0.859
AE1a	0.781	0.098	0.621	0.614	-1.165 / -0.071 / 0.830
AE1b	0.884	0.077	0.703	0.506	-1.206 / -0.237 / 0.668
AE3	0.744	0.089	0.592	0.650	-1.453 / -0.552 / 0.258
<i>Factor 2</i>					
AA2	1.000		0.598	0.642	-1.088 / -0.643
FA2	1.280	0.211	0.766	0.413	-2.138 / -0.618 / 0.720
FA3	1.028	0.186	0.615	0.621	-1.743 / -0.431 / 0.643
FA5	1.119	0.194	0.669	0.552	-1.845 / -0.774 / 0.830
FB1	0.883	0.172	0.529	0.721	-2.138 / -0.830 / 0.618
FB2	1.194	0.210	0.714	0.490	-1.743 / -0.643 / 0.343
FE2	1.044	0.178	0.625	0.610	-1.743 / -0.830 / 0.386
FE3	1.196	0.208	0.716	0.488	-1.845 / -0.802 / 0.546
FE4	1.206	0.212	0.721	0.480	-1.657 / -0.920 / 0.694
FF1	0.749	0.138	0.448	0.799	-1.206 / -0.343 / 0.694
<i>Factor 3</i>					
AD1	1.000		0.523	0.727	-1.291 / -0.386 / 0.774
FD1	1.279	0.202	0.669	0.553	-1.581 / -0.694 / 0.596
FD2	1.903	0.313	0.995	0.009	-1.344 / -0.386 / 0.774
<i>Factor 4</i>					
FA4	1.000		0.638	0.592	-1.165 / -0.092 / 1.017
FC1	1.285	0.495	0.820	0.327	-0.859 / -0.010 / 0.830

*Proportion of variance accounted for my measurement error

Table 5.14: Parameter estimates for Model 1 – continued from previous page

Factor	Unstandardised variance/covariance	SE	Standardised variance/covariance
<u>Factor 1</u>	<u>0.633</u>	<u>0.083</u>	<u>1.000</u>
- Factor 2	0.358	0.071	0.753
- Factor 3	0.266	0.050	0.639
- Factor 4	0.168	0.072	0.331
<u>Factor 2</u>	<u>0.358</u>	<u>0.112</u>	<u>1.000</u>
- Factor 3	0.229	0.047	0.731
- Factor 4	0.108	0.054	0.281
<u>Factor 3</u>	<u>0.273</u>	<u>0.078</u>	<u>1.000</u>
- Factor 4	0.036	0.036	0.107
<u>Factor 4</u>	<u>0.408</u>	<u>0.170</u>	<u>1.000</u>

Table 5.15: parameter estimates for Model 2

Parameter	Factor Loadings (Unstandardised)	SE	Factor Loadings (Standardised)	ME Variances	Thresholds
<i>Factor 1</i>					
AA2	1.000	0.000	0.625	0.610	-1.093 / -0.649
AB1	1.259	0.200	0.787	0.381	-1.849 / -0.957 / 0.245
AB2	1.083	0.194	0.676	0.543	-1.585 / -0.625 / 0.394
AB3	0.988	0.191	0.617	0.620	-1.585 / -0.460 / 0.674
AC1	1.068	0.193	0.667	0.555	-1.747 / -0.416 / 0.865
AE1a	0.991	0.188	0.619	0.617	-1.170 / -0.081 / 0.836
AE1b	1.087	0.182	0.679	0.539	-1.170 / -0.224 / 0.674
AE3	0.947	0.179	0.592	0.650	-1.457 / -0.529 / 0.245
<i>Factor 2</i>					
FA2	1.000	0.000	0.757	0.426	-2.141 / -0.625 / 0.700
FA3	0.797	0.079	0.604	0.635	-1.747 / -0.438 / 0.625
FA5	0.891	0.084	0.675	0.544	-1.849 / -0.780 / 0.836
FB1	0.703	0.083	0.532	0.717	-2.141 / -0.636 / 0.600
FB2	0.931	0.101	0.705	0.503	-1.747 / -0.649 / 0.329
FE2	0.816	0.090	0.618	0.618	-1.747 / -0.836 / 0.394
FE3	0.932	0.087	0.706	0.502	-1.849 / -0.808 / 0.552
FE4	0.945	0.085	0.716	0.488	-1.661 / -0.895 / 0.700
<i>Factor 3</i>					
FD1	1.000	0.000	0.655	0.571	-1.585 / -0.700 / 0.552
FD2	1.478	0.146	0.968	0.062	-1.349 / -0.394 / 0.753
<i>Factor 4</i>					
FA4	1.000	0.000	0.645	0.585	-1.170 / -0.081 / 1.023
FC1	1.264	0.537	0.815	0.336	-0.865 / 0.000 / 0.836

Factor	Unstandardised variance/covariance	SE	Standardised variance/covariance
<u>Factor 1</u>	<u>0.390</u>	<u>0.123</u>	<u>1.000</u>
- Factor 2	0.366	0.066	0.774
- Factor 3	0.268	0.051	0.654
- Factor 4	0.136	0.064	0.338
<u>Factor 2</u>	<u>0.574</u>	<u>0.080</u>	<u>1.000</u>
- Factor 3	0.383	0.046	0.773
- Factor 4	0.111	0.061	0.226
<u>Factor 3</u>	<u>0.429</u>	<u>0.063</u>	<u>1.000</u>
- Factor 4	0.029	0.044	0.068
<u>Factor 4</u>	<u>0.415</u>	<u>0.189</u>	<u>1.000</u>

As becomes clear from Table 5.13, both models provided a suitable fit to the data, with slightly preferable results (e.g., a higher X^2 p-value and a lower RMSEA) for Model 2. Variances as a result due to measurement error are above the ideal threshold of .500 for both models: that is, more than half of the variance in ratings of most indicators (items) is not explained by the model but instead by other factors (Kline, 2011). This implies that although these models provide the only acceptable fit with the data out of all compared models that suggested a potential fit based on theory or the previously conducted EFA, neither model explains the majority of the variance found in the data.

- Confirmed factor structure

Although it is important to make use of SEMs that fit a dataset used, the real goal is to test the validity of a theory by specifying models that make use of appropriate indicators to measure predicted relationships between plausible constructs within that theory (Hayduk et al., 2007; Kline, 2011). In other words: a model should not only fit the data but also make sense theoretically. Earlier in this section, it became clear that of the two models that passed the appropriate fit tests, Model 2 provided preferable fit statistics. In addition, this model is best in line with the theory on the basis of which the confirmatory factor analysis was carried out, that states that the activities and functions of project-based learning in Studio Schools load onto separate factors when measured through implementation questionnaires completed by students. Therefore, Model 2 was chosen as the factor structure to be used in further analyses that include PBL implementation data. Figure 5.3a depicts this model when specified at the individual participant level, with all original indicator names (e.g., AA2 for Activity A2) left intact. Figure 5.3b depicts the model with the revised indicator names that will be used further for the purpose of clarity. Some further explanation of Figures

5.3a and 5.3b should also be provided. As can be seen, the activity factor and three function factors that emerged within Model 2 have been given names in line with the manifest indicators they relate to. All activities and the vast majority of functions load onto single factors, and were therefore labelled ‘Core PBL Activities’ and ‘Core PBL Functions’. The two other function factors related to more specific manifest indicators and where named in line with these. Indicators F21 and F22 (originally labelled FD1 and FD2) were concerned with students learning their CREATE-skills (non-cognitive skills prioritised in Studio Schools), hence the factor associated with these was labelled ‘CREATE-skills development’. Indicators F31 and F32 (originally FA4 and FC1) were originally hypothesised to cluster with other variables, but loaded best together. Both related to the perceived relevance by students of the subject-based curriculum they learn as their projects. Therefore, function factor three was labelled ‘Perceived relevance of curriculum’.

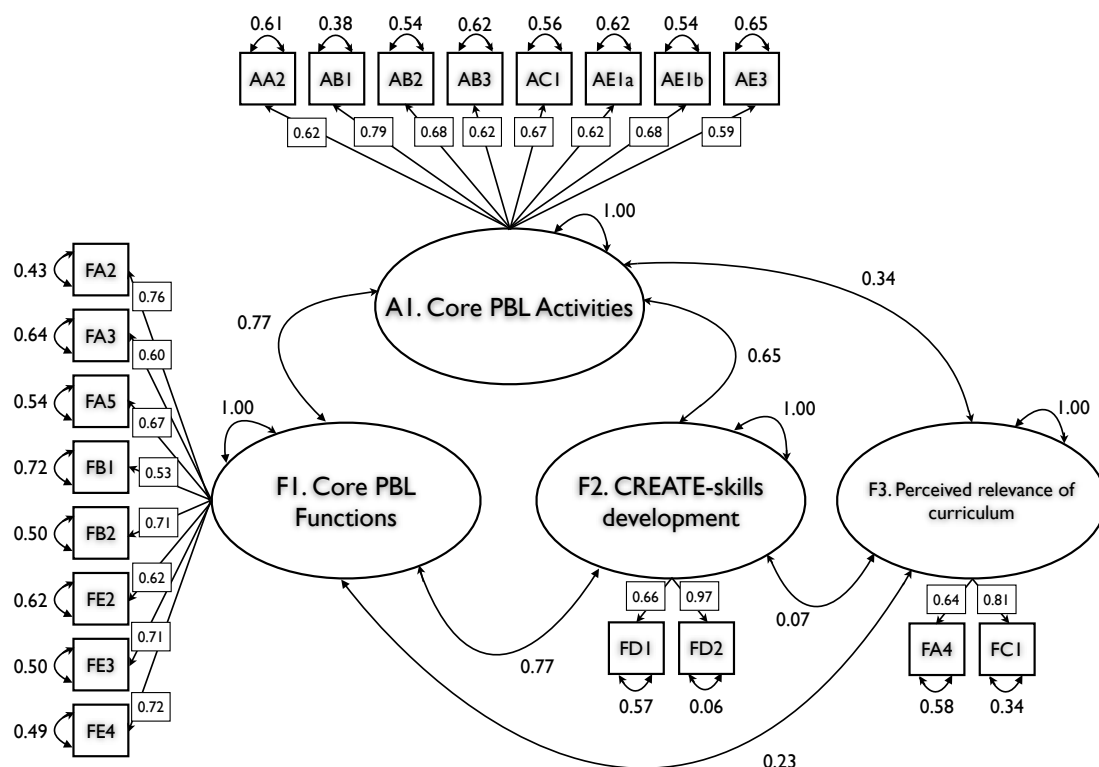


Figure 5.3a: Overview of factor structure used to analyse PBL implementation data, estimated using WLS estimation procedures, using initial indicator names

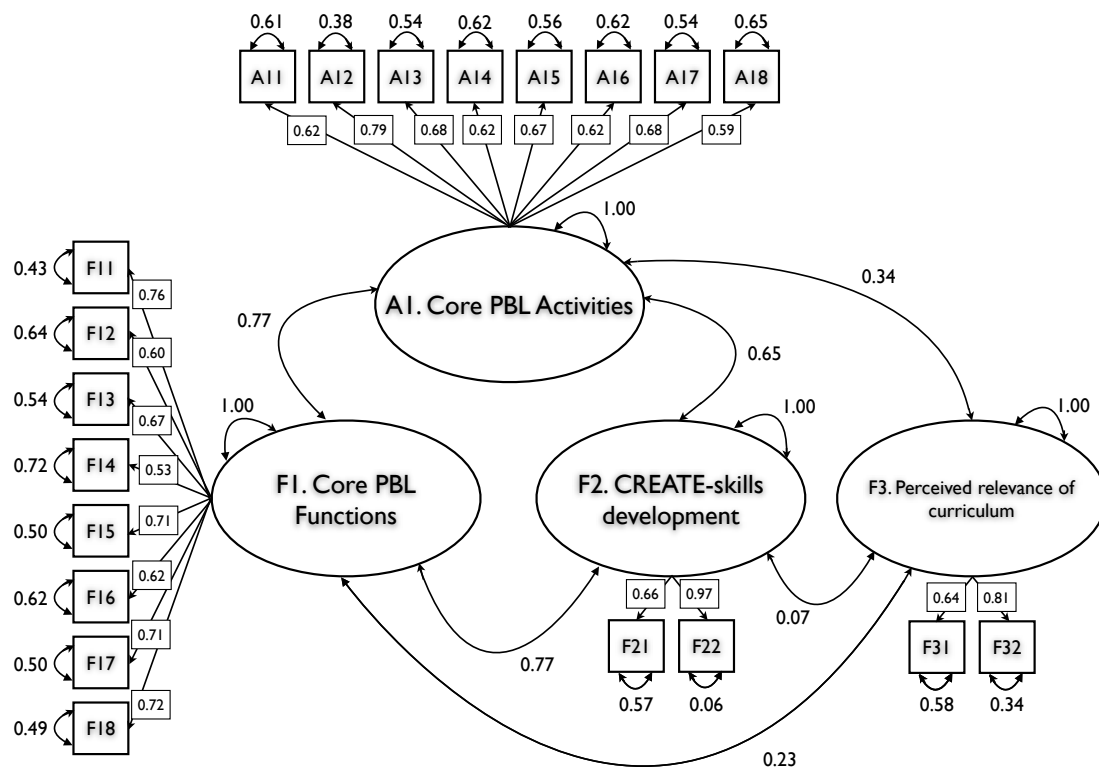


Figure 5.3b: Overview of factor structure used to analyse PBL implementation data, estimated using WLS estimation procedures, using revised indicator names

Figures 5.3a and 5.3b display the high amount of variance that is explained as a result of measurement error for many items that was also noted in Table 5.15. The figures further suggest that the implementation (as seen by students) of PBL activities strongly covaries with two function factors, namely the Core PBL Functions (F11-18) and the CREATE-skills development function (F21-22), and somewhat with the third function factor, namely students' perceived relevance of their curriculum (F31-32). The Core PBL Functions also correlate strongly with the CREATE-skills development function, but neither of these correlate in particular with the Perceived relevance of curriculum function. Although very high or very low covariance between latent variables suggests poor divergent or convergent validity respectively, the nature of the variables and their relationships analysed here requires a slightly different approach to the appraisal of these statistics.

Repeated assessment of internal consistency

The original PBL implementation measures described earlier in this paragraph were revised based on the results of the exploratory and confirmatory factor analyses of student-reported PBL implementation data described in the previous paragraphs (see Table 5.10 for the item recoding/removal process, and Table 5.15. for the revised factor structure). Upon completion of this process, the revised measure was analysed for its internal consistency once more. The results of this analysis are presented in Table 5.16.

Table 5.16: Internal consistency statistics for revised scales and subscales used to monitor project-based learning in Studio Schools upon completion of factor analyses

Scale	α (N items) - Teacher	α (N items) - Students	α (N items) - Researcher
Core PBL Activities	.892 (8)	.803 (8)	.711 (8)
Core PBL Functions	.745 (5)	.808 (8)	.156 (4)
CREATE-skills development	.573 (2)	.686 (2)	X* (0)
Perceived relevance of curriculum	.589 (2)	.580 (2)	.501 (3)
All Activities	.871 (9**)	.789 (9**)	.747 (9**)
All functions	.800 (9)	.802 (12)	.539 (7)
Overall	.898 (18)	.880 (21)	.763 (16)

* No α 's were calculated for this scale

** The activity originally labelled 'AD1' was excluded from the Core PBL Activities factor because it did not load well with its items but was included in the model because it correlated with function factor two ('CREATE-skills development'). This explains the one item difference between the Core PBL Activities and the overall set of activities

As Table 5.16 demonstrates, the activities and function scales as specified under the factor structure that was confirmed on the basis of student-reported data show good internal consistency when completed by teachers and students. When completed by the researcher, the activities scale also shows good internal consistency, but the internal consistency of the function scale is poor. This suggests a difference in interpretation of the constituent items between the researcher and participants.

Personal coaching

Initial assessment of internal consistency

As for the project-based learning measures, the internal consistency of the original measures for measuring the implementation of personal coaching was assessed (see paragraph 4.4.3). The results of this analysis are presented in Table 5.17.

Table 5.17: Internal consistency statistics for original scales and subscales used to monitor personal coaching in Studio Schools

Scale	α (N items) - Coach	α (N items) - Students	α (N items) - Researcher
AA	.718 (7)	.530 (7)	.453 (7)
FA	.894 (14)	.864 (14)	.786 (14)
AB	.191 (3)	.642 (3)	-.014 (3)
FB	.423 (3)	.472 (3)	.463 (4)
AC	.286 (2)	.022 (2)	.295 (2)
FC	.639 (2)	.447 (2)	X* (1)
AD	X* (1)	X* (1)	X* (1)
FD	.775 (4)	.871 (4)	X* (1)
All Activities	.756 (13)	.490 (13)	.424 (13)
All functions	.910 (23)	.901 (23)	.792 (20)
Overall	.912 (36)	.884 (36)	.770 (33)

* No α 's were calculated for scales comprising a single item

Similarly to the PBL instruments, the presented α 's for the activity scales used for the student- and researcher-reported coaching measures presented in Table 5.17 are rather poor. The overall function scales are consistent for all raters, but some of the constituent subscales are not. Thus, in line with the original PBL scales, some items that form part of the original personal coaching scales appeared inconsistent with other items.

Assessment of polychoric correlations among items

As discussed in paragraph 5.3.1, the sample sizes for the data collected from students (N=35 sessions), coaches (N=36 sessions), and the data completed by the researcher (N=36 sessions) did not allow for most factor analysis procedures. Therefore, a further assessment of the underlying factor structure of the collected data was done through inspection of the polychoric correlations amongst items within each dataset

(see Appendix C3). This revealed similar problems to those presented by inspection of the tables generated for project-based learning: some items correlated weakly or significantly negatively with other items. The correlation structures were also somewhat different for the different raters, suggesting again a difference in interpretation of the constituent items that will be discussed further in section 5.4. Table 5.18 presents items that were removed from the scales based on an assessment of their correlations with other items. In one case, activity AD1 was removed upon a repeated inspection of the internal consistency of coaching scales after the other items presented in Table 5.18 were removed. The implications for analysis that result from the removal of items are presented in paragraph 5.6.2 and in Chapter VI.

Table 5.18: removal of personal coaching items poorly correlated with overall structure in student questionnaire

Original item name	Action	Reason
AA1b	Removed	Not significantly related to any other item (highest correlation = 0.42), with substantial negative correlations
AA4b	Removed	Not significantly related to any other item (highest correlation = 0.46), with substantial negative correlations
AB3	Removed	Not significantly related to any other item (highest correlation = 0.48) with substantial negative correlations
AC2	Removed	Not significantly related to any other item (highest correlation = 0.47) with substantial negative correlations
FC1a	Removed	Negative correlations (highest correlation = -.054) with majority of items
AD1	Removed	Negative correlations with most items in researcher scale, poor consistency with all activities scale

Repeated assessment of internal consistency

Once the revisions presented in Table 5.18 were made to the original personal coaching scales, a repeated analysis of the internal consistency was conducted for the revised scales. Table 5.19 presents the results of this analysis.

Table 5.19: Internal consistency statistics for revised scales and subscales used to monitor personal coaching in Studio Schools

Scale	α (N items) - Coach	α (N items) - Students	α (N items) - Researcher
AA	.739 (5)	.469 (5)	.545 (5)
FA	.894 (14)	.864 (14)	.786 (14)
AB	.563 (2)	.889 (2)	-.375 (2)
FB	.423 (3)	.472 (3)	.463 (4)
AC	X* (1)	X* (1)	X* (1)
FC	X* (1)	X* (1)	X* (1)
AD	X* (0)	X* (0)	X* (0)
FD	.775 (4)	.871 (4)	X* (1)
All Activities	.791 (8)	.513 (8)	.536 (8)
All functions	.919 (22)	.883 (22)	.792 (20)
Overall	.924 (30)	.893 (30)	.793 (28)

* No α 's were calculated for scales comprising one or less items

As becomes clear from Table 5.19, the overall function scales show good internal consistency across all raters. The activities scale shows good internal consistency when completed by coaches, but much less consistency when completed by students or researchers. This provides further evidence for interpretation differences between the rater groups. These differences will be further examined in the next section.

5.4 Inter-rater agreement

5.4.1 Methods and measures

The data used for assessing the inter-rater agreement between the researcher and participants for the used measures were the same data that were collected for the assessment of the internal consistency and construct validity presented in the previous paragraph. As noted before, these data were collected using paper-based quantitative questionnaires items were rated on ordinal scales by the researcher and participants respectively. Sessions were sampled as per the sampling design outlined in paragraph 5.2.1.

Ratings by additional researchers

In addition to ratings obtained from participants and the doctoral researcher, two additional researchers rated three project-based learning sessions –plus one additional session for piloting– (one additional researcher per session) based on in-vivo observations using the same instrument used by the doctoral researcher. Another additional researcher rated six personal coaching (plus three additional sessions for piloting) sessions based on recordings of the session, again using the same instrument used by the doctoral researcher. All additional researchers hold Master’s degrees in a social science discipline. The first four coaching sessions to be rated by the additional researcher were selected randomly per school so that at least one session per school would be independently rated. The final two sessions were selected randomly from the remaining sessions across all schools. PBL and coaching sessions that were used for piloting purposes were sampled specifically for these purposes and not included in any further analyses. Because the independent rating by an additional researcher of PBL sessions required the physical presence of the researcher in the classroom, these were sampled in deliberation with the schools. The three double-rated sessions took place in three different schools.

Measures used for determining inter-rater reliability

Intra-class correlation coefficients were calculated across all items included in the questionnaires before and after revisions were made to the used scales to assess inter-rater agreement between researchers as well as between the doctoral researchers and participants, and between the two participant groups. The ICC forms a measure of inter-rater reliability through assessing the extent to which ratings by two or more raters belonging to the same class (in this case ratings by two raters of the same session) are correlated. Its value can range between zero and one, with a number

closer to one representing higher reliability. The ICC was selected as the measure for determining reliability because all rating scales were ordinal (a suboptimal condition for the use of Kappa statistics) and multiple raters (i.e. teachers and groups of students) rated different PBL and coaching sessions (a suboptimal condition for simple correlation measures). Table 5.20 provides an overview of the analysis procedures used to assess the inter-rater agreement between raters for project-based learning and personal coaching.

Table 5.20: Overview of analysis procedures used to determine the inter-rater reliability of used measures

Component of focus	Analysis	Data analysed
Project-based learning	<i>Initial assessment of inter-rater agreement (Intra-class Correlation Coefficients)</i>	- Student implementation questionnaires - Teacher implementation questionnaires - Researcher implementation questionnaires - Additional researcher implementation questionnaires
	Revision of scales	
	<i>Repeated assessment of inter-rater agreement</i>	- Student implementation questionnaires - Teacher implementation questionnaires - Researcher implementation questionnaires - Additional researcher implementation questionnaires
Personal coaching	<i>Initial assessment of inter-rater agreement</i>	- Student implementation questionnaires - Teacher implementation questionnaires - Researcher implementation questionnaires - Additional researcher implementation questionnaires
	Revision of scales	
	<i>Repeated assessment of inter-rater agreement</i>	- Student implementation questionnaires - Coach implementation questionnaires - Researcher implementation questionnaires - Additional researcher implementation questionnaires

5.4.2 Procedure

The inter-rater reliability of the used implementation measures was evaluated twice using the methods and measures described in the previous paragraph. The first

assessment was done across all items originally included in the questionnaires. After the measures were revised based on the findings of analyses relating to internal consistency and construct validity, a repeated assessment was conducted across all remaining items.

5.4.3 Findings

Project-based learning

The results of the inter-rater agreement analyses between all rater groups for the original as well as revised measures used for monitoring project-based learning are presented in Table 5.21. The table shows strong inter-rater agreement between researchers, both before and after revision of the used measure. This suggests that the used instrument is reliable when used by independent observers with a background in research. Although there appears to be some amount of agreement between researchers and teachers, the amount is too small to support reliability. For the final measure, that was selected as the preferred option based on its internal consistency and construct validity, inter-rater agreement between students and teacher as well as the researcher is very poor. Table 5.16 in paragraph 5.3.3 further showed different levels of measure internal consistency when the measure is used by students and teachers versus the researcher. This suggests a different interpretation between different rater groups with regards to the PBL implementation practices in Studio Schools when rating according to the used measure, as well as perhaps a different interpretation of the meaning of some of the included items. Implications of these findings are discussed in paragraph 5.6.5.

Table 5.21: Inter-rater agreement statistics for original and revised measures used to monitor project-based learning in Studio Schools

Inter-rater agreement (ICC) under original factor structure			
<u>Between Researchers</u>	<u>Researcher - Students</u>	<u>Researcher - Teachers</u>	<u>Teachers - Students</u>
.898	.111	.411	.588
Inter-rater agreement (ICC) under revised factor structure			
.892	.239	.472	.278

Personal coaching

Table 5.22 presents the results of the inter-rater agreement analyses between the rater groups for the personal coaching measure. This table presents a pattern of agreement similar to that of the project-based learning measure, with high reliability between researchers and much lower reliability between the different rater groups, although reliability between teachers and students is moderate and much higher than for project-based learning. Combined with the findings presented in Table 5.19 regarding different levels of internal consistency of the personal coaching measure for different rater groups, these findings suggest that –similar to PBL– there are substantial interpretation differences between the rater groups when rating the implementation of personal coaching in Studio Schools.

Table 5.22: Inter-rater agreement statistics for original and revised measures used to monitor personal coaching in Studio Schools

Inter-rater agreement (ICC) under original factor structure			
<u>Between Researchers</u>	<u>Researcher - Students</u>	<u>Researcher - Teachers</u>	<u>Teachers - Students</u>
.888	.078	.118	.578
Inter-rater agreement (ICC) under revised factor structure			
.898	.090	.387	.600

5.5. Overview of definitive measures

5.5.1 Alterations to original measures based on psychometric assessment

Paragraph 4.4.3 of the Chapter 4 provided an overview of the measures used for monitoring the implementation of PBL and personal coaching in Studio Schools as they were originally developed based on theory and empirical piloting. The previous sections of the current chapter reported the iterative assessments of the psychometric properties of these measures and revisions made to the measures based on the results of these assessments. This paragraph provides a summary of what was changed to the measures as compared to the original measures in paragraph 4.4.3. The next

paragraph provides a tabulated overview of the definitive measures used for determining implementation fidelity of PBL and coaching.

Project-based learning

The internal consistency of the original PBL activities and functions scales was suboptimal. Exploratory factor analyses then led to the removal of two items and the reverse coding of three. Further exploratory and confirmatory factor analyses suggested the student data supported a single activity factor plus one separate item instead of multiple activity scales, and three instead of six function factors. Some items that were not hypothesised to load together loaded better together than with the scales as part of which they were expected to load in theory. The internal consistency of the revised measure, particularly the combined activities and functions scales, was much stronger than that of the original measure. It needs to be noted here that, as items were removed and recoded and revisions were made to the construction of the activities and functions scales, the Studio School model elements (PBL and coaching) and change processes of which the implementation will be reported in the next chapter are somewhat different from the ones that were originally specified in Chapter IV. In other words, the theory of change that is measured with the definitive measure is somewhat different from the theory on which the original measure is based. The implications of this are discussed in paragraph 5.6.3.

Personal coaching

Reliability analyses demonstrated that the internal consistency of some subscales as these were originally specified was poor. Exploratory factor analyses suggested this was partly caused by some items that loaded poorly with the rest of the scales, and these items were removed. Because the internal consistency of some subscales was

still suboptimal after removing these items, the decision was made to combine all activities and functions in single scales for each respective construct. This decision necessarily acknowledges that linear relationships between individual activities and their functions (as specified in theory) are difficult to measure validly for personal coaching in Studio Schools.

5.5.2 Definitive measures for monitoring implementation of project-based learning

Table 5.23: Overview of definitive scales and items used for measuring implementation of project-based learning in Studio Schools

Scale	Items
<i>A1. Core PBL Activities</i>	A11. The problems or assignments central to project activities should be presented with structured guidance on where to start in order to complete it A12. When assisting students, learning coaches should present solutions to encountered problems A13. When assisting students, learning coaches should instruct students what they need to study to complete the project successfully A14. Students should draw upon key methods of enquiry suitable with the project's discipline(s) (e.g., document analysis for English/History, collecting own data for Science) A15. Students should draw on content of one or multiple subject(s) that are offered at Studio Schools (i.e. English language, English literature, math, science) whilst completing the project A16. Students should receive explicit oral or written feedback/critique on the products (deliverables) of their project activities from their classmates, as well as provide it to them A17. Students should receive explicit oral or written feedback/critique on the products (deliverables) of their project activities from their teacher A18. Classroom time should be spent as much as possible on working with subject content (independently developing/applying knowledge) and as little as possible on lecturing subject content
<i>A2. CREATE-skills development</i>	A19. During the process of working on the project, students should reflect explicitly on which of their CREATE skills they need to further develop to successfully complete the project
<i>F1. Core PBL Functions</i>	F11. The problem central to the project should be of importance to students

	F12. It should be clear to students and teacher how studied subject content aligns with the end product(s) of the project F13. The adults involved in the project should inspire students to put in their best effort F14. Students should discover independently where/how the subject content and other factual information that is needed to understand/manage a problem and complete the project can be obtained F15. Students should feel ownership of the project and its end product(s); it is their responsibility to come with the answers to complete it successfully F16. Students recognise those aspects of their own and others' work that can/should be improved F17. Students recognise those aspects of their own and others' work that are particularly strong F18. Students develop their own best way of learning/studying subject content for it to be of use towards project tasks and end products
<i>F2. CREATE-skills development</i>	F21. It should be clear to a student and teacher which CREATE skills a student needs to develop further to complete the project successfully F22. It should be clear to a student and learning coach how students' project activities help them develop <u>specific</u> CREATE skills
<i>F3. Perceived relevance of curriculum</i>	F31. Students' application of knowledge and skills reflects that of adult life F32. It should be clear to students and teachers how content/knowledge that is needed to complete the project links to the subject content they require to learn to pass their exams

5.5.3 Definitive measures for monitoring implementation of personal coaching

Table 5.24: Overview of definitive scales and items used for measuring implementation of personal coaching in Studio Schools

Scale	Items
<i>A. Personal Coaching Activities</i>	A1. Student and coach should reflect on student's CREATE progress A2. Student and coach should reflect on student's behaviour A3. Student and coach should discuss the cues and situations that negatively affect student's performance at school/work A4. Student and coach should discuss the cues and situations that positively affect student's performance at school/work A5. Coach should guide student towards discovering methods for reinforcing constructive thought patterns and emotions A6. Student and coach should discuss and determine concrete curricular and personal goals of the student

	A7. Student and coach should discuss and determine the concrete actions that need to be undertaken to achieve these goals A8. During coaching sessions, coaches should use questions to help students self-examine their behaviours and attitudes A9. During coaching sessions, coaches should not offer concrete advice or direct students towards solutions or actions
<i>F. Personal Coaching Functions</i>	F1. After a coaching session, coach should be aware of students' specific strengths with regards to their academic and CREATE progress F2. After a coaching session, student should be aware of students' specific strengths with regards to their academic and CREATE progress F3. After a coaching session, coach should be aware of students' specific weaknesses with regards to their academic and CREATE progress F4. After a coaching session, student should be aware of students' specific weaknesses with regards to their academic and CREATE progress F5. After a coaching session, coach should be aware of students' specific strengths with regards to their behaviour F6. After a coaching session, student should be aware of students' specific strengths with regards to their behaviour F7. After a coaching session, coach should be aware of students' specific weaknesses with regards to their behaviour F8. After a coaching session, student should be aware of students' specific weaknesses with regards to their behaviour F9. After a coaching session, coach should be aware of the thought patterns that influence students' perceptions of themselves and others negatively F10. After a coaching session, student should be aware of the thought patterns that influence students' perceptions of themselves and others negatively F11. After a coaching session, coach should be aware of the thought patterns that influence students' perceptions of themselves and others positively F12. After a coaching session, student should be aware of the thought patterns that influence students' perceptions of themselves and others positively F13. After a coaching session, students should be aware of specific methods for reinforcing constructive thought patterns and emotions F14. After a coaching session, students should be aware of specific methods for reshaping dysfunctional thought patterns and emotions F15. It should be clear to the student/personal coach and rating researcher what the curricular and personal goals are that the student is working towards F16. It should be clear to students/personal coach and rating researcher what actions are to be undertaken by i) the student and ii) the coach to achieve these goals.

	F17. After each coaching session, student/coach and rating researchers should feel that specified changes that are needed to improve student's behaviours/situation/learning were defined primarily by the student. F18. Coach should trust student F19. Student should trust coach F20. Student and coach should feel that coach is aware of and understands students' personal characteristics F21. Student should feel supported by coach when engaging with new activities or (interactive) strategies with which they may be uncomfortable
--	--

5.6 Discussion

This chapter reported the findings of a key step towards monitoring function-focused implementation fidelity in Studio Schools, namely a psychometric assessment of the instruments used in the study and alterations made to these as part of the process.

These findings are further discussed in this section.

5.6.1 Choice of method for answering this research question

In this study, the decision was made to measure implementation in Studio Schools through quantitative questionnaires, which was therefore naturally also the method of choice to examine the validity and reliability of a function-focused approach to measuring fidelity. A measurement structural equation model was then used to analyse the data collected with these questionnaires in order to answer the second research question of the thesis. This decision was made to allow for a data collection plan that was feasible within the scope of a doctoral project, but is not free of limitations. Although triangulation of sources in terms of implementation observers (i.e., the researcher, students, teachers and coaches) was used to gain various perspectives, only the student data were used in a confirmatory factor analysis of the PBL instrument, because these yielded the only sample size needed at a minimum for this technique. Thus, a multitrait-multimethod (MTMM) analysis was not possible in

this study, impeding the ruling out of methodological issues explaining the patterns found in the data (Dishion et al., 1999). This presents a risk of ‘monomethod bias’ with regards to the findings presented in this chapter (Cook & Campbell, 1979). It also makes it impossible to determine if a method problem, as defined by Bank and colleagues, exists: highest correlation between indicators that are derived from the same measurement method (Bank et al., 1990). Confirmatory factor analysis was selected as the method of choice for assessing the empirical relationship between activities and functions, because it is the most rigorous method for determining the empirically supported structure of measurement instruments. Nevertheless, the findings presented here must be interpreted with some caution.

5.6.2 Operational issues faced during data collection

As mentioned in paragraph 5.2.1, the original plan for sampling PBL and coaching sessions and participants for the project contained a much larger number of sessions to be sampled, and the intention to sample sessions and dates randomly. Although the Studio Schools Trust and the management of the participating schools approved this sampling plan, it proved impossible to adhere to it during the process of collecting data in the schools as a result of some operational issues faced during the project. The doctoral candidate recognises the benefit of facing and dealing with such challenges early on in an academic career, and some reflection on how these were handled in the project and could have been handled in an optimal situation is provided below.

School capacity issues

All four schools operated under seriously limited capacity, in the sense that they all worked with staff numbers representing (close to) the minimum necessary to function. This is a direct result of a friction between a school model that intends to include a

maximum of 300 students per school to allow for small class sizes and a personal approach to education, and a national per-pupil education-funding model. Limited capacity by the schools implied it was impossible to randomly select days and sessions for data collection, as these were necessarily selected based on staff availability. Staff sickness repeatedly resulted in last-minute cancellation of data collection visits, and could not always be compensated during the remainder of the project. Furthermore, it created an apparent reluctance by the school study coordinators to plan visits in a structured manner, because they could not foresee their capacity on a given day two weeks or more prior to that day. Ideally, a larger number of schools would have been sampled for the study, as to spread the burden of visits. However, this was not possible given the geographic spread of Studio Schools across England and the availability of only one researcher (the doctoral candidate) to collect the data.

Large interest in Studio Schools from other parties

Studio Schools are a new model of education that aims for schools to be thoroughly involved in their local communities. Naturally, this creates an interest from an overwhelming number of parties that aim to collaborate with, study, or help improve the schools. During the planning process of visits for data collection, one reason that was often given for cancellation or reluctance to plan visits was the need for a school to focus on visits by other parties. Such parties included OFSTED, businesses interested in providing work places, and community organisations providing opportunities for real world PBL. Because such organisations are core to the everyday operation of Studio Schools, it follows that the study reported in this thesis did not always have the schools' priority. Although this was expected beforehand, the doctoral candidate did not anticipate the magnitude of this issue. Better

communication (i.e. discussing the issue explicitly, including a strategy for handling it during the project) at the onset of the study might have prevented conflicting schedules from interfering with the process of data collection to the extent that they did.

5.6.3 Modification of theory during the psychometric assessment process

The original measures reported in paragraph 4.4.3 of the previous chapter were based on a programme theory that was specified in detail during multiple stages, using multiple sources including the scientific literature and stakeholder input. These measures were used to monitor implementation of activities and the presence of resulting hypothesised change mechanisms (functions) in Studio Schools. As described in this chapter, some items were removed from the measures or reverse coded to improve the psychometric properties of the latter upon completion of data collection. This implies that the Studio School model elements (PBL and coaching) and change processes of which the implementation will be reported in Chapter 6 (which uses the revised measures) are somewhat different from the ones that were originally specified in Chapter 4. It is recommended that, upon psychometric assessment, mismatching items should be removed from scales or recoded depending on the size and their direction with other items (DeVellis, 2003). It is also common to revise programme theories upon their empirical evaluation (Connell & Kubisch, 1998; Weiss, 2004). In this study, significant and strong negative correlations were found for three PBL items during exploratory factor analysis. All three items represented measures of the extent to which PBL teachers gave directions to students versus supporting their independent learning (item AA2 relates to the amount of direction given to students when their project tasks are presented, and items AB1 and AB2 relate to directions given to students during the PBL instruction process). Although a

lack of direction given to students while working on their project assignments formed an explicit part of the original programme theory of PBL in Studio Schools, this hypothesis was reversed during the measurement development process as a result of the recoding of these items. This implies a revision of the programme theory of PBL in Studio Schools that assumes that, on the basis of empirical data collected in four schools, a large amount of student versus teacher directedness may not actually form part of the programme theory of Studio Schools PBL as it is seen and implemented by teachers and students. Considering the fact that PBL was originally designed for undergraduate education (Barrows, 1996), it is possible that this aspect of its original model may not be well-suited –at least not without some additional components and staff training– for education at the secondary level (Ertmer et al., 2005; Kirschner et al., 2006; Kolodner et al., 2003). Anecdotal evidence from classroom observations suggests this might be the case, as it was observed regularly and in some cases confirmed by teachers that students (younger year groups in particular) had some difficulties working independently without direction. However, the possibility of other factors lying at the root of the mismatch between items AA2, AB1, and AB2 as they were originally coded –such as, for example, suboptimal phrasing of the questions by the researcher or misinterpretation of these by students or responses subject to social desirability bias– cannot be ruled out. Informed by the suspicion that the relevant items as originally coded mismatched the overall instruments, the Studio Schools PBL theory of change was revised in order to allow for a sounder instrument from a psychometric perspective that is based on an adapted programme theory. Preference was thus given to using instruments with relatively stronger psychometric properties informed by an assessment of student-reported implementation data over instruments that were based on an original programme theory. Ideally, separate data –

involving development samples— would have been collected for the purpose of exploratory factor analysis, upon which data would be collected for confirmatory factor analysis, and finally for the purpose of evaluating implementation in Studio Schools upon validation of the instruments (DeVellis, 2003). In addition, valid outcome data would have aided in evaluating the validity of the programme theories on which the instruments were based. However, these options were not available within the study for reasons explained earlier.

5.6.4 PBL confirmatory factor analysis results further examined

Low levels of variance accounted for

The measurement model that was specified to confirm the factor structure for the PBL implementation ratings provided by the students demonstrated the best theoretical and empirical fit. However, more than half of the variance in ratings of most manifest indicators was not explained by the model but instead by other factors. There may be several reasons for this. First, it is theoretically possible that the model specified an erroneous factor structure based on the available data. This is to say that there is a possibility that the specification of different latent variables on which manifest indicators would load in different combinations could, in theory, result in a better explanation of the variance. However, the specification and comparison of various measurement models to the preferred model, each based on a plausible theoretical framework and/or exploratory factor analysis results but each providing suboptimal fit, largely rules out this possibility (Bowen & Guo, 2011; Kline 2011). A second factor that may lie at the root of the limited explained variance is the possibility that important manifest indicators (i.e. items) were not measured that are in fact relevant to the theory of change of PBL in Studio Schools, either because they load on the set of specified activity or function factors or on an entirely different factor or set of

factors that was not measured. Although the PBL theory of change was carefully specified through an iterative process of examination of the scientific literature and consultation with stakeholders, the possibility that an inaccurate PBL programme theory lays at the root of the low observed levels of explained variance cannot be ruled out. Finally, the limited variance explained by the model could be caused by the fact that factors related to the completers of the questionnaires (i.e. the students) explain the majority of variance found in the data. For example, it is possible that factors such as a student's level of engagement, concentration or understanding of the asked questions while completing the questionnaire is a better predictor of the ratings s/he gives than the specified model. The overall low levels of variance found in the data provided by some students suggest that this explanation is plausible, although it is impossible to provide a definitive cause for the limited variance explained by the specified measurement model.

Further reflection on the confirmed factor structure

The results of the confirmatory factor analysis of the student PBL data revealed a factor structure in which activities and functions loaded onto separate, yet highly covarying, factors. This pattern is unusual and requires further discussion. First, The high covariance between the activity factor and the first two function factors ('Core PBL functions' and 'CREATE skills development') could indicate limited divergent validity of the measure. However, it needs to be noted that it is hypothesised that activities *cause* their functions and it would therefore be surprising if these showed little or no covariance. The nature of the relationship between activities and functions therefore changes the interpretation of these results, because very high levels of covariance between such factors are expected, whereas this is normally not preferred in factor analyses. Given these results, a key task was the investigation of alternative

factor structures, particularly those in which activities and functions were allowed to load onto the same factors. This is because if such models were to be admissible, a different theory of the empirical relationship between activities and functions –one that argues that both concepts can/should be integrated into the same factor as part of an implementation measure– would become more plausible. To investigate this relationship further, multiple CFA models were compared in which manifest indicators loading onto strongly covarying factors were specified to load together onto single factors, which led exclusively to inadmissible solutions (see Appendix C4). Thus, it is concluded here that the high levels of covariance found between these factors do not warrant aggregating of the factors into a single factor. The question remains what these results mean for the theoretical concept of a programme component comprising a (set of) activity/activities and a (set of) function(s) that is introduced in this thesis, and how it should be measured. In this study, the decision was made to hypothesise and treat activities and functions as indicators loading onto separate factors, but as part of the same measure. This approach was supported by the factor analysis results, but more inquiry is needed to offer conclusive support for the theorized relationship –one of exclusivity but also of covariance– between both concepts that is proposed as part of this thesis. More thought, examples of operationalisation, and discussion of the challenges involved in these will be needed to improve our understanding of these concepts.

Second, whereas high levels of covariance were found for the three factors discussed above, very low levels of covariance were found for function factor 3 ('Perceived relevance of the curriculum') versus the other two function factors, and a low level was found between the activity factor and function factor 3. This suggest that FF3 may not be triggered solely by project-based learning, but that other elements of the

Studio Schools model might also affect it, which is plausible. However, the possibility that other factors such as those discussed in the previous subparagraph may have affected the relationship between the different factors specified in the model cannot be ruled out.

Finally, although the confirmed factor structure provided an optimal (empirical and theoretical) fit to the data as it was specified at the individual level, it lost its fit when accounting for clustering of students' responses within the sessions and schools that they were provided. This was a result of a sample size that proved underpowered to allow for a SEM that accounted for clustering (all compared models lost their fit when accounting for clustering). Accounting for clustering is important because a failure to do so may result in inaccurate model fit evaluation as well as parameter estimates (Pornprasertmanit et al., 2014) if clustering amongst responses is high. An item-by-item as well as scale-by-scale analysis of the intra-cluster correlation coefficients (ICCs) with respects to clustering of the data within schools as well as within sessions revealed that this was not the case. It was decided to continue the analyses at the individual level in this thesis so that the process of testing and validating a measure based on a function-focused approach towards fidelity could be demonstrated and some preliminary conclusions could be provided with regards to the potential validity of such a measure. Nevertheless, the lack of accounting for clustering in the confirmed factor structure further warrants a cautious approach towards interpretation of the results of this study.

5.6.5 Inter-rater agreement results further examined

Inter-rater agreement between the different rater groups for the definitive measures was poor overall, ranging from .239 (students and the researcher) to .472 (teachers

and the researcher) for project-based learning and .090 (students and the researcher) to .600 (students and coaches) for personal coaching. Between researchers, inter-rater agreement was very strong with an ICC of .892 for project-based learning and .898 for personal coaching. All ICCs improved upon revising the measures, except for the relationship between PBL ratings given by the students and the teachers. Differences between the ratings supplied by different informants are common in the literature (Basch et al., 1985; Breitenstein et al., 2010; Lillehoj et al., 2004). Some possible explanations can be discussed for low levels of inter-rater agreement in this study. First, it is possible that the different rater groups provided different ratings because they focused on different aspects of project-based learning and personal coaching while providing their ratings. For example, because students work in small groups in PBL, there is the possibility that students and teachers saw a different implementation than the researcher and/or one another. Although this seems less likely for personal coaching (a one-to-one activity), selective recall bias by the students and coaches may –as for PBL– still have affected their ratings as they rated sessions after these were completed whilst the researcher completed his checklist while watching a recording that could be stopped and rewound if necessary (Breitenstein et al., 2010; Pereplechikova et al., 2007). Second, the different groups of raters may have observed the exact same implementation but interpreted what they saw differently. The differences in levels of ratings between rater groups discussed in the next chapter provide some evidence for this, with much lower ratings for the researcher than for the teachers and students. Although not necessarily with this magnitude, previous multi-informant implementation studies have found similar results (Basch et al., 1985; Lillehoj et al., 2004). Third, there is a chance that the different rater groups interpreted the questions asked from them in their checklists differently based on

different interpretations of the models of PBL and coaching they were rating. For example, when asked to rate if a teacher instructed students what to study in order to complete a project (PBL) or if a coach came up with most of the solutions to a student's problems (coaching), it is possible that the different raters had different interpretations of what these questions meant and how their answers would add up to a 'good' implementation score. Some strong differences found between the rater groups with regards to the patterns of internal consistency found in their ratings (see paragraph 5.3.3) provide some evidence for this explanation. No explanation for low inter-rater agreement can be confirmed or ruled out based on the collected data and results found. Additionally, in the absence of outcome data, no evidence of which ratings best predict outcomes can be provided. Therefore, the fidelity ratings of all three rater groups will be reported, discussed, and taken into account in the next chapter.

5.6.6 Sampling issues

The decisions made in this study regarding the sampling of schools, sessions, and participants for data collection affect the conclusions that can be drawn from the data in some important ways.

Sampling strategy

All samples were selected based on convenience. This decision was made bearing in mind the formative nature of the study as well as limitations with regards to resources faced by the researcher as well as the participating schools. However, this causes some limitations regarding the generalisability of the results reported in this chapter, which are discussed in this paragraph.

Sample sizes

In the end, a much smaller sample of sessions (hence participants) was recruited than intended. The most frequent reasons for this were long intervals between receiving responses from study coordinators, their reluctance to plan visits due to an overload of visits from other parties with an interest in Studio Schools (including inspecting bodies), and last-minute cancellation of visits due to understaffing as a result of staff illness. As sessions were selected non-randomly, this small sample size further adds to the limited generalisability of these results. Additionally, it inhibited accounting for clustering in the specified CFA models without automatically losing all fit as a result of low statistical power. For an overview and discussion of the sample sizes and characteristics of each participating school, the reader is referred to section 6.2. of the next chapter.

Using students as the sample of focus for confirmatory factor analysis

In order to assess the feasibility and validity incorporating a function-focused approach into fidelity monitoring, it was decided a psychometric assessment of the instruments used was key. The only sample that allowed for a factor analysis, an integral part of this assessment, was the sample of PBL data provided by the students. In addition, because student outcomes are the core outcomes of the Studio School model and no previous validation of the Studio Schools theories of change for PBL and coaching had been conducted, it was decided student-reported data would be the appropriate data source of focus at this point. Nevertheless, based on the different patterns found regarding the internal consistency of the used scales depending on which rater group data were provided by, there is a good chance that the definitive instruments used as well as conclusions regarding their validity would have looked somewhat different had other raters been the sample of focus for these analyses.

Because the student PBL sample was still on the small side for a confirmatory factor analysis (DeVellis, 2003), the generalisability of the CFA is limited further.

5.6.6 Further implications for function-focused fidelity monitoring

In Box 3.1 (paragraph 3.3.3, Chapter III), three underlying assumptions of a function-focused approach towards implementation fidelity monitoring were noted that should be discussed in light of the results of this study.

1. Occurring functions are the result of implementation of an activity or set of activities, and the extent to which the presence of specified functions correlates to the implementation of the activity/activities needs to be demonstrated empirically.
2. Provided that all functions of all implemented activities are specified appropriately in a valid theory of change, the relationship between implemented activities and outcomes is mediated in its entirety by the functions of the activities
3. The extent to which specified functions are related to specified outputs and outcomes corresponds to the validity of programme theory, and not to the specific form of an implemented set of programme activities

Box 3.1 (repeated): Three underlying assumptions of a function-focused approach towards defining and monitoring implementation fidelity

The empirical demonstration of a correlation between functions and form (assumption 1 in Box 3.1) will be addressed in the next chapter, and will not be discussed here. However, the results discussed in the current chapter demonstrate that specifying and evaluating a valid and very detailed programme theory in which the outcomes of implemented activities are mediated in entirety by the functions of those activities (assumptions 2 and 3) is easier said than done. Some items in the used activity and function scales for PBL and coaching needed recoding or removal because of their misfit. Although the internal consistency of the aggregated form and function scales was good, the consistency of some subscales (i.e. components of PBL and coaching as defined by form and function) was poor, and differed strongly by rater group. This demonstrates that it is difficult to get a detailed, micro-level theory of change exactly

right, and raises the question if an approach based on linear assumptions between form and functions offers the best use of both concepts. It has been argued before that researchers typically do not know which components of a programme theory may be responsible for which effects (Mihalic, 2004). Additionally, researchers recognise that the effects of a set of specific intervention components is often greater than the sum of the effects of their parts (Bonell et al., 2012) and cannot be attributed solely to them but are instead affected by other factors as well (Chen & Rossi, 1983; Pawson & Tilley, 1997). Furthermore, some acknowledge that functions are more difficult to specify because knowledge about the links between early and proximal outcomes from scientific enquiry as well as evaluation experience is not well developed (Connell & Kubisch, 1998). This implies that careful consideration should be given to specifying theories of change accurately at the micro-level, and that multiple revisions based on an iterative process of iterative implementation monitoring and stakeholder consultation may be needed to ‘get a model right’. In some cases, the interaction between a programme’s components may be too complex for linear modelling between activities and functions to be useful, and an approach in which aggregate scales are used, as they ultimately are in this study, might be preferable for the sake of preserving validity of measurement.

5.7 Conclusion

Can the implementation of complex social interventions be measured validly and reliably in terms of the functions, next to the form, of their components?

To measure implementation fidelity in this project, instruments were produced with scales comprising exclusively of sets of activities or functions, and that had

acceptable-to-very-strong internal consistency. A confirmatory factor analysis of student-reported ratings of fidelity to project-based learning resulted in an optimal solution in which activities and functions loaded onto separate factors. The selected model ‘passed’ all fit tests in spite of a very small sample size, although it did not explain the majority of the variance found in the data, and did not account for clustering. Inter-rater agreement was suboptimal between different rater groups, but high agreement between researchers suggests it can be feasible to rate fidelity of function alongside fidelity of form. Study limitations inhibit the drawing of definitive conclusions. Specifically, limitations related to the sampling strategy employed to collect the data for psychometric assessments limit the generalisability of the results to other PBL and coaching sessions in the participating schools and beyond. Furthermore, in studies with a longer time-span and that are subject to a large amount of resources, instruments like these that demonstrate some suboptimal psychometric properties would be improved and tested again on an independent sample. It therefore needs to be emphasised that the results reported here serve a demonstrative purpose (i.e. to demonstrate how a function-focused approach towards implementation fidelity could be designed, executed and validated). This study demonstrated that functions and form can be specified and measured simultaneously for complex interventions, through the use of methods commonly used in process evaluations. The used measures were evaluated and showed properties that support the hypothesis that it is possible to measure functions and form validly and reliably as long as instruments are carefully designed and validated repeatedly in line with best practices in psychometrics. The study further indicated that a micro-level approach in which individual or small clusters of activities and functions form subscales as part of larger scales may be very difficult to execute with validity and reliability, and that careful

thought should be given to the process of modelling the interaction between activities and functions if this approach is to be of benefit to evaluations.

5.8 Chapter summary

In this chapter, the validity and reliability of the measures developed in Chapter IV were assessed. Data on the implementation of project-based learning and personal coaching sessions were collected in four participating Studio Schools during a four-month period. Exploratory factor analyses as well as an analysis of the reliability (internal consistency and inter-rater) were conducted, upon which the measures were revised. The revised measures were assessed again for their psychometric properties through repetition of the reliability analyses as well as a confirmatory factor analysis for the student PBL instrument. These revised measures were used to analyse the implementation practices of the participating schools by form and function, the process and results of which will be reported in the next chapter. This study further showed that functions can be measured alongside form as part of a measure of implementation fidelity, and that the instruments used for doing so can be assessed for their validity and reliability in line with common psychometric practice.

Chapter VI: Current state of implementation in Studio Schools

6.1 Introduction

The previous chapters (4 and 5) described the development of the implementation measures used in this project that are based on a novel conceptualisation of implementation fidelity. In the current chapter, the results of the implementation study that was carried out in Studio Schools on the basis of these measures are presented.

6.1.1 Research questions

This chapter addresses the third and fourth research questions of the doctoral project:

- *Comparing across schools, does the Studio School model currently appear to be implemented with fidelity defined by form and function?*
- *To what extent are form-focused and function-focused fidelity scores for project-based learning and personal coaching in Studio Schools correlated?*

Based on the relationship between form and function that was outlined in Chapter III (i.e. function may be more standard than form in flexible complex interventions), it was hypothesised that functional fidelity would be rated higher on average than fidelity of form, and that smaller differences in scores across schools would be found. Additionally, a substantial –but less than perfect– amount of correlation would be expected between the scores, because form and function strive towards the same outcomes but may not always co-occur. These hypotheses will be revisited in this chapter, in light of its findings.

6.2 Methods

As noted at the beginning of section 5.2 of the previous chapter, the same data that were collected to assess the validity and reliability of the used instruments were also used to analyse implementation practices in Studio Schools. Hence, the methods used to collect implementation data are summarized below to avoid repetition, and the reader is referred to Chapter 5 for a detailed overview.

6.2.1 Settings and sampling

The full process of sampling for this study is described in paragraph 5.2.1 of the previous chapter. A short summary of the sampling strategy is provided here.

Sampling of schools, sessions, and participants

Four Studio Schools participated in the study, and were selected through convenience sampling in consultation with the Studio Schools Trust. Twelve sessions of project-based learning and 36 personal coaching sessions were included in the study. These sessions were also sampled on the basis of convenience with the assistance of principals and the teachers and coaches responsible for facilitating the respective sessions. Written informed consent was obtained from all participating teaching staff and students. For the convenience of the reader, Tables 5.1 and 5.4 of the previous chapter, here labelled Tables 6.1 and 6.2 respectively are included once more below.

Table 6.1: Overview of sessions, students, and teachers included per school in PBL data collection (Copied from Table 5.1)

School	1	2	3	4	Total
PBL sessions (N)	3	4	3	2	12
Students – observations (N)	66	24	25	30	145
Students – unique (N)	45	18	25	30	118
Teachers – observations (N)	3	4	3	2	12
Teachers – unique (N)	2	2	3	2	9

Table 6.2: Overview of sessions, students, and coaches included per school in coaching data collection (Copied from Table 5.4)

School	1	2	3	4	Total
Coaching sessions (N)	12	14	6	4	36
Students – observations (N)	12	13*	6	4	35
Students – unique (N)	12	13	6	4	35
Teachers – observations (N)	12	14	6	4	36
Teachers – unique (N)	3	5	4	2	14

*One student did not complete the questionnaire

6.2.2 Measures

Implementation measures

The development process of the used implementation measures has been described in Chapter IV, and is summarized in Figure 6.1. Chapter V reported on the validity and reliability of these measures and described further modifications that were made. This process is summarized in Figure 6.2. An item-by-item overview of the definitive measures used in the study is also provided in paragraphs 5.5.2 (PBL) and 5.5.3 (coaching) of the previous chapter. All implementation data were collected using paper-based quantitative questionnaires in which participants (teaching staff and students) and the researcher rated items (i.e. statements and questions) regarding the implementation of activities and the occurrence of functions. Students and teachers also answered a series of qualitative questions that provided the data used by the researcher to rate those items of the fidelity scales that could not be observed in vivo. The questionnaires used by the different rater groups (i.e. learning coaches and personal coaches, learners, and the researcher) mostly contained the same questions and statements, although some questions and statements were only asked of those who could provide a valid answer. For example, the researcher did not provide any rating for the statement that ‘learners felt ownership of their projects’ (one of the items included in the PBL checklist) because it was judged impossible for the

researcher to validly rate this statement. Please refer to Appendix B for the questionnaires used.

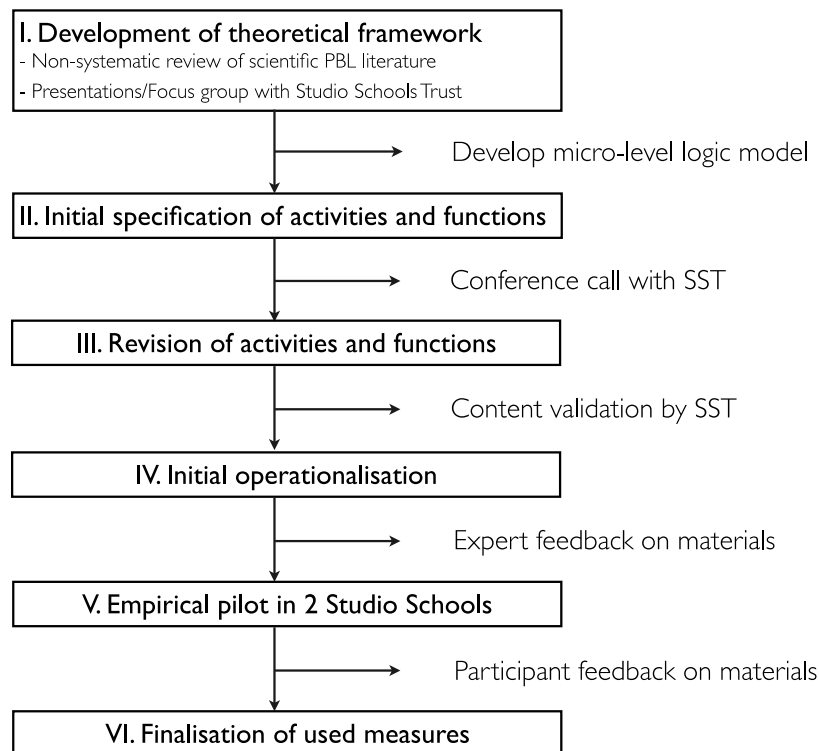


Figure 6.1: Flowchart depicting the development of implementation measures for Studio Schools (for a full overview of the development process, see Chapter 4)

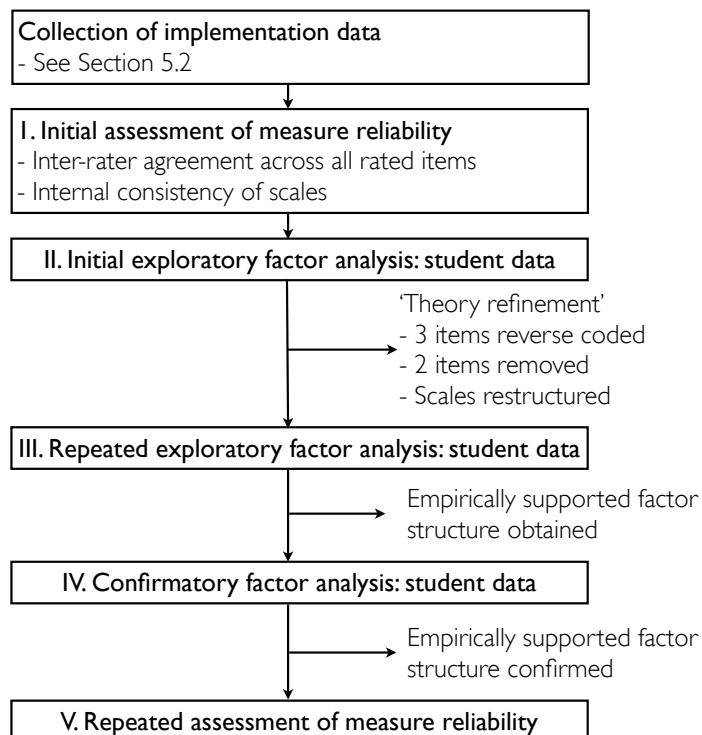


Figure 6.2: Flowchart depicting analysis procedure for determining the validity and internal consistency of used measures (copied from Figure 5.1)

Numeric properties of used scales

- Project-based learning

As previously presented in paragraph 5.5.2, the project-based learning activities (i.e. form-focused) scale comprised of a subscale representing ‘core PBL activities’ and an additional item representing ‘CREATE skills development’. The function scale contained three subscales, representing ‘core PBL functions’, ‘CREATE skills development functions’, and ‘perceived relevance of curriculum functions’ respectively. Items were rated on ordinal scales containing either the options ‘Yes’ (2) or ‘No’ (0), or a 4-option scale ranging from ‘Completely disagree’ (0) to ‘Completely agree’ (3). Table 6.3 presents the numeric properties of the PBL activities and functions subscales as well as the overall scales. The different numbers of items and corresponding maximum scores on the function scales for each rater group reflect the fact that not all questions could be rated by each rater group. For example, fewer ratings could be given by the teachers and researcher on scale F1 compared to the students, and no ratings could be given by the researcher on scale F2.

Table 6.3: Numeric properties of PBL scales, shown per rater group

Scale	N unique items	Max score teachers (N items)	Max score students (N items)	Max score researcher (N items)
A1	8	23 (8)	23 (8)	23 (8)
A2	1	3 (1)	3 (1)	3 (1)
<u>Activities</u>	<u>9</u>	<u>26 (9)</u>	<u>26 (9)</u>	<u>26 (9)</u>
F1	8	15 (5)	24 (8)	12 (4)
F2	2	6 (2)	6 (2)	0 (0)
F3	2	6 (2)	6 (2)	9 (3)
<u>Functions</u>	<u>12</u>	<u>27 (9)</u>	<u>36 (12)</u>	<u>21 (7)</u>

- Personal coaching

Because psychometric assessment of the scale suggested some internal consistency issues for several subscales, the decision was made to aggregate the activities and

functions of personal coaching into one single scale for each (see paragraph 5.5.1), labelled ‘core personal coaching activities’ and ‘core personal coaching functions’. Table 6.4 presents the numeric properties for these scales.

Table 6.4: Numeric properties of personal coaching scales, shown per rater group

Scale	N unique items	Max score researcher (N items)	Max score students (N items)	Max score teachers (N items)
Activities	8	17 (8)	17 (8)	17 (8)
Functions	21	60 (20)	66 (22)	66 (22)

6.2.3 Procedure

All PBL researcher questionnaires were completed in the classrooms of the participating schools based on in vivo observations during PBL lessons. Students and teachers completed their questionnaires at the end of the session they had participated in. Qualitative responses submitted by teachers and students were coded by the researcher after completing the data collection visit. Personal coaching sessions were not attended in vivo by the researcher but instead recorded and coded upon completing the data collection visit. Teachers and students completed their questionnaires during the data collection visits. For a more detailed overview of the procedures for collecting implementation data on PBL and coaching, the reader is referred to paragraph 5.2.2 of the previous chapter.

6.2.4 Data analysis

Implementation data

Coding of qualitative responses

Because some activities and functions could not be observed by the researcher, teachers and students were asked to provide qualitative answers to a series of questions on the questionnaires they completed at the end of PBL and coaching

sessions. Examples of the questions asked include ‘please describe the subjects you have had to use today while working on the project’ and ‘please describe this student’s strengths and weaknesses with regards to their academic and CREATE progress’. Teachers were also asked if they had made any adaptations to lessons or coaching sessions and, if so, why they had made these, and it was originally planned to code these data quantitatively to assess how different adaptations and the reasons for these related to fidelity of form and function (see Moore et al., 2013 for an example). However, not enough data were provided to allow for a meaningful quantitative analysis of these data, and the data lacked the richness for a useful qualitative analysis. Hence, these data were discarded. Responses were coded quantitatively by the researcher on an ordinal scale containing either the options ‘Yes’ (2) or ‘No’ (0), or a 4-option scale ranging from ‘Completely disagree’ (0) to ‘Completely agree’ (3), in line with the rest of the researcher implementation checklist. A selected sample of sessions was coded by two researchers for the purpose of determining inter-rater reliability of the coding process (see paragraph 5.4.3).

Calculating implementation scores

Implementation scores were calculated per session per rater group for each subscale (if applicable) by adding the raw scores given to individual items belonging to that subscale. For student PBL ratings, averages per PBL lesson were calculated. Average scores per school and across all schools were then calculated per subscale for each rater group. Subscale scores were combined into an overall score for each activity (i.e. form-focused) and function (i.e. function-focused) scale, and average scores as well as their range were calculated from these scores for each school and the overall sample. All scores were subsequently divided by the maximum score for each scale or subscale in order to express levels of implementation fidelity in the participating

schools as a percentage of maximum fidelity. All implementation scores were calculated in SPSS version 22.

Analysis of the relationship between form-focused and function-focused fidelity scores

In order to assess the extent to which form-focused and function-focused fidelity of PBL and personal coaching as rated in Studio Schools are correlated, correlation coefficients were calculated per rater group for the aggregate form-focused and function-focused fidelity scores. Spearman's ρ (as opposed to Pearson's r) was used as a measure for correlation because the small samples per session and the organisation of fidelity ratings on an ordinal scale violate the assumption that the implementation data are normally distributed. All correlation analyses were conducted in SPSS version 22.

6.3 Results

6.3.1 Student sample characteristics

At the beginning of the 2014-2015 school year (September 2014), all participating students were asked to complete a questionnaire with regards to their demographic and academic backgrounds. These data were collected in order to put the current state of implementation in Studio Schools into context, and to assess if the ethos of contextual adaptation within the Studio Schools movement is warranted based on the diversity of its student populations. In total, 209 students enrolled at the four participating schools completed the survey. Students had to have started their Studio School education at the beginning of the 2014-2015 academic year in order to be eligible for participation. A group of 239 students were eligible, of which 30 (12.6%) were not included because they were absent on the days of data collection or because

they did not provide their written consent for participation. As becomes clear from Table 6.5, students who completed the background survey were enrolled exclusively in Years 10 and 12. The majority of participants were enrolled in Year 10 at the time of the baseline questionnaire. In addition, the number of included students is not distributed equally across the participating schools. These differences reflect the different sizes and compositions of the student bodies within the participating schools. Schools 01 and 02 are Studio Schools that are not directly linked to any other school, whereas Schools 03 and 04 are directly linked to a supporting Academy. Therefore, they are more likely to relocate students from these Academies to complete the Sixth Form at their new Studio Schools if this is deemed suitable, rather than mostly enrolling Year 10 students from feeder schools as is the case for Schools 01 and 02. Because only a very small fraction of Year 12 students in Schools 01 and 02 had joined their school for the first time at the start the 2014-2015 school year, it was decided to focus only on Year 10 students in these schools to make efficient use of resources.

Table 6.5: Number of students who completed the baseline questionnaire

School	1	2	3	4	Total
N	51	24	52	82	209
% of eligible participants	83.6	88.9	94.5	85.4	87.4

Demographics

Studio Schools aim to attract students with backgrounds representative of the general population (Studio Schools Trust, 2012a). To assess if this is the case for the included sample of students, they were asked to report on their demographic backgrounds. These data are reported in Table 6.6. As can be inferred from the table, the overall mean age of participating students was 15.06 years. This figure is in line with the

dominance of Year 10 students in the sample, completed with a substantial group of Year 12 students. As shown in Table 2, the mean age of the included sample of students differs somewhat between each school. This is in line with the unequal distribution across different year groups between the schools. Overall, more girls (59.8%) than boys (42.6%) participated in the study, although more boys than girls participated in Schools 01 and 02. Although several factors (e.g., the industry of specialisation of the school) might lie at the root of this, an examination of these was beyond the focus of this study.

Table 6.6: Overview of students' demographic backgrounds

School	1	2	3	4	Total
Age					
Mean	14.49	14.43	15.67	15.22	15.06
SD	.32	.29	1.11	1.01	.98
Gender					
Girls (%)	45.1	45.8	71.2	59.8	57.4
Year group (N)					
10 (N)	51	24	22	54	151
12 (N)	0	0	30	28	58
Ethnicity (%)					
White British	89.8	91.7	84.3	82.7	85.9
Black or Black British	4.1	0.0	3.9	8.6	5.4
Mixed ethnicity	2.0	0.0	5.9	2.5	2.9
Other	4.1	8.3	5.9	6.2	5.8
Academic ability					
Mean teacher assessment level at Year 9	5.94	5.56	5.20	5.94	5.73
SD	.95	.81	1.11	.67	.91
Household composition					
Lives w both biological parents (%)	69.4	37.5	49.0	60.5	57.1
Biological parent and stepparent (%)	18.4	20.8	31.4	19.7	22.4
Single parent and/or other adults (%)	12.2	41.7	19.6	19.7	20.5
Socioeconomic indicators					
<i>Education completed by parents</i>					
No GCSEs (%)	7.4	6.3	0.0	7.9	5.9
GCSEs only (%)	37.0	37.5	44.8	47.6	43.7
GCSEs + A-levels (%)	22.2	6.3	13.8	22.2	18.5
Undergraduate Degree	3.7	31.3	17.2	12.7	14.1
<i>Employment by parents</i>					
Any adult in household employed (%)	95.6	91.3	88.5	85.0	89.0
<i>Family affluence</i>					
Mean Family Affluence Scale (FASC)	6.51	6.62	5.54	6.00	6.08
SD	1.65	1.41	1.91	1.88	1.82

The vast majority of the students that completed the baseline questionnaire were of White British origin (85.9%). The second largest group of participating students were Black or Black British (5.4%). The ethnic composition of the included group of students is roughly in line with the national average of students enrolled in secondary schools (DfE, 2015), with a slight overrepresentation of students who were White British (compared to a national average of 78.0%) and an underrepresentation of Asian or British Asian students (2.0% compared a national average of 9.2%). Because the current study represents the only research conducted into Studio Schools, it is currently unclear how the included student bodies compare to the overall bodies of the participating Studio Schools as well as other Studio Schools. The large percentage of included students suggests that they are likely a good representation of the students that were eligible for participation in the participating schools. Differences regarding the ethnic composition of the student bodies between schools may be caused by the areas in which the schools are located. More than half (57.1%) of the included group of students currently live with both their biological parents, although substantial differences exist between the schools. Somewhat over a fifth (22.4%) live with a biological parent and a stepparent, leaving roughly a fifth (20.5%) to live with a single parent or with (an)other adult(s). There were also important differences between the participating schools with regards to the fraction of guardians who had completed higher education. For example, whereas 43.8% of students in School 02 lived with at least one adult with an Undergraduate degree, these numbers were much lower for School 01 (7.4%), School 03 (24.1%), and school 04 (19.0%). Students' household employment rates varied between schools from 85.0% to 95.6%. These rates are well above the average Employment Rate in England (71.7%) as well as these rates for the areas in which each school is located (ONS, 2014b). However, the

statistics can only offer a broad indication as employment was measured at the household level within this survey, rather than at the individual level (at which it was measured to compile national and regional averages). There is currently no accurate indication that the students who participated in the study came from households that were substantially more or less employed than the average household in England.

Socioeconomic status as measured by the Family Affluence Scale (FASC)

Adolescent-reported data on their parents' education and employment is not always reliable, as they are often unable to accurately recall these. As a solution to this problem, the *Family Affluence Scale* was developed. The FASC asks some simple material affluence questions that are easy to answer by adolescents (see Box 1). Research has demonstrated that FASC reports by adolescents show a good match with family socioeconomic status measures reported by parents (Currie et al., 2008).

Box 1: The Family Affluence Scale (Source: Currie et al., 2008)

1. Does your family own a car, van or truck?
2. Do you have your own bedroom for yourself?
3. During the past 12 months, how many times did you travel away on holiday with your family?
4. How many computers does your family own?

Scores on the FASC can be analysed using three point categories, with a score of 0-2 indicating low affluence, scores of 3-5 indicating middle affluence, and 6-9 indicating high affluence at the international level. In 2006, researchers reported a national average for England of 5.3, indicating middle- to high affluence (Boyce et al., 2006). The students included in the current survey reported an overall score of 6.08, indicating relatively high affluence (see Table 6.6). Nevertheless, the difference between the FASC scores reported by students in this survey and the national average of 2006 are not substantial, and it is possible that the national average has increased over time. Thus, there is no grounded reason to assume that the FASC scores of the

included students are very different from the national average. Small differences existed between Schools 01, 02 and 04, with students from School 03 reporting statistically significantly lower affluence.

Academic ability

To test if the participating Studio Schools attract students of all ability levels, the participating students were asked at baseline to report their Teacher Assessment levels at the end of Year 9. In the case of School 03, additional TA-level data were obtained from School records. Overall, data were obtained for 171 students. On average, the included students obtained a TA level of 5.73 at the end of Year 9, which is in line with the national average (UK Government, 2014). The average TA levels of students enrolled at School 03 were significantly lower than the averages of other schools, although there is a chance that this could be explained by the fact that School 03 provided data for non-completers whereas the other schools did not provide these data. The absolute averages of all schools were within the range of one level (see Table 6.6).

Motivation

Studio Schools offer an approach to education that is distinguishable from mainstream secondary education, offering several features (such as project-based learning, personal coaching, and realistic work experience) that are different from what is taught at conventional academies. To determine which of these were of particular interest to the participating students, they were asked to indicate their most important reasons for choosing to continue their education at a Studio School. Table 6.7 presents the results of this question, based on the responses of 196 students (97.6%). As becomes clear from the table, the most important reason for choosing a Studio School according to the participating students is to learn and study in a small school

environment, followed by the opportunity to gain work experience. The majority of students further indicated that learning things that are important outside of school and getting the necessary support from their teachers and peers were also amongst their most important reasons for choosing a Studio School. Differences in expectations between the populations of different Studio Schools require different pedagogic foci to be successful.

Conclusion

Although the participating schools overall appear to have attracted representative student bodies, marked differences between the populations of the different schools in terms of their demographic background and expectations suggest that the Studio Schools ethos of tailoring the model to specific school circumstances may indeed be necessary to make it effective at the national level.

Table 6.7: Most important reasons for choosing a Studio School, as reported by students who completed the baseline questionnaire (N=196)

School	1	2	3	4	Total
Most important reason for choosing a Studio School (% 'Yes' per reason)					
Learning and studying in a small school environment	84.4	77.3	71.4	62.5	71.4
Getting work experience	62.2	63.6	65.3	77.5	69.4
Learning things that are important outside of school	46.7	40.9	65.3	63.7	57.7
Getting the support needed from others at school	46.7	54.5	59.2	58.8	55.6
Guidance and coaching by a Personal Coach	37.8	50.0	44.9	46.3	44.4
Following own personalised curriculum	40.0	45.5	38.8	36.3	38.8
Project-Based Learning	33.3	45.5	20.4	38.8	33.7
Dislike of other school options	26.7	31.8	30.6	8.8	20.9
Other reasons	2.2	0.0	18.4	5.0	7.1

6.3.2 Implementation of project-based learning in Studio Schools

Scores per rater group

Tables 6.8 and 6.9 present the fidelity scores and percentages as defined by form and functions respectively, as rated by the different rater groups in observed PBL lessons.

Teacher-rated fidelity

Across all schools, teachers rated the fidelity of their PBL activities with a score of 17.16, corresponding to a percentage of 66% (range 35%–100%) of the maximum score. Very large differences exist between the average ratings of teachers within and between the schools, with scores on the activities scale ranging from a score representing 38.5% of the maximum score as rated by teachers in School 01 versus 98.1% in School 04, and scores ranging up to 30% in School 02. The presence of PBL functions was rated 19.17, corresponding to a slightly higher fidelity percentage of 71 (range 37%–100%). Between-school differences in reported functional fidelity are smaller than those reported for form, but still show a substantial range between 64.2% (as rated in School 01) and 98.2% (as rated in School 04). This result is in line with the hypothesis that function-focused fidelity scores may be somewhat higher and consistent across sites for a complex intervention like Studio Schools, although session-level (i.e. within-school) differences were slightly higher.

Table 6.8: Project-based learning fidelity scores (%) as defined by form, shown per rater group

School	1	2	3	4	Total
Number of sessions observed	3	4	3	2*	<u>12</u>
Avg. ratings by teachers					
Overall	10.00 (38.5)	14.25 (54.8)	22.63 (87.0)	25.50 (98.1)	17.16 (66.0)
(Overall range – percentages)	(35 -42)	(35 – 65)	(81 – 100)	(96 – 100)	(35 – 100)
A1	9.00 (39.1)	14.25 (62.0)	20.33 (88.4)	23.00 (100.0)	15.92 (69.2)
A2	1.00 (33.3)	0.00 (0.0)	2.29 (76.4)	2.50 (83.3)	1.24 (41.3)
Avg. ratings by students					
Overall	15.89 (61.1)	17.41 (67.0)	18.36 (70.6)	19.10 (73.5)	17.55 (67.5)
(Overall range – percentages)	(56 -70)	(59 – 73)	(57 – 79)	(67 – 80)	(56 – 80)
A1	14.12 (61.4)	15.85 (68.9)	16.49 (71.7)	17.21 (74.8)	15.80 (68.7)
A2	1.78 (59.2)	1.56 (52.1)	1.87 (62.4)	1.89 (63.0)	1.75 (58.3)
Avg. ratings by researcher					
Overall	5.33 (20.5)	7.25 (27.9)	12.67 (48.7)	15.00 (57.7)	9.42 (36.2)
(Overall range – percentages)	(12 - 38)	(19 – 38)	(38 – 46)	(42 – 73)	(12 – 73)
A1	5.33 (23.2)	7.25 (31.5)	11.67 (50.7)	14.50 (63.0)	9.08 (39.5)
A2	0.00 (0.0)	0.00 (0.0)	1.00 (33.3)	0.50 (16.7)	0.33 (11.1)

Table 6.9: Project-based learning fidelity scores (%) as defined by function, shown per rater group

School	1	2	3	4	Total
Number of sessions observed	3	4	3	2*	<u>12</u>
Avg. ratings by teachers					
Overall	17.33 (64.2)	15.50 (57.4)	21.00 (77.8)	26.50 (98.2)	19.17 (71.0)
(Overall range – percentages)	(56 – 78)	(37 – 67)	(67 – 96)	(96 – 100)	(37 – 100)
F1	9.67 (64.4)	8.50 (56.7)	12.00 (80.0)	15.00 (100.0)	10.75 (71.7)
F2	5.33 (89.9)	2.75 (45.8)	4.00 (66.7)	6.00 (100.0)	4.25 (70.8)
F3	2.33 (39.9)	4.25 (70.8)	5.00 (83.3)	5.50 (91.7)	4.17 (69.4)
Avg. ratings by students					
Overall	21.88 (60.1)	23.88 (66.3)	24.12 (67.0)	24.37 (69.6)	22.95 (65.7)
(Overall range – percentages)	(54 – 67)	(62 – 76)	(62 – 75)	(65 – 74)	(54 – 76)
F1	15.46 (64.4)	16.83 (70.1)	16.76 (69.8)	17.71 (73.8)	16.62 (69.2)
F2	3.56 (59.4)	4.35 (72.6)	3.89 (64.8)	3.81 (63.6)	3.95 (65.8)
F3	2.86 (47.7)	2.69 (44.9)	3.46 (57.7)	3.53 (58.9)	3.07 (51.1)
Avg. ratings by researcher					
Overall	5.78 (27.5)	7.75 (36.9)	12.33 (58.7)	8.17 (38.9)	8.47 (40.4)
(Overall range – percentages)	(10 – 44)	(29 – 52)	(52 – 62)	(24 – 55)	(10 – 62)
F1	4.39 (36.6)	4.50 (37.5)	7.33 (61.1)	4.34 (36.1)	5.15 (42.9)
F2	X	X	X	X	X
F3	1.39 (15.4)	3.25 (36.1)	5.00 (55.6)	3.84 (42.6)	3.32 (36.9)

Student-rated fidelity

On average across observed sessions in all participating schools, student-rated fidelity of form was 17.55, representing a percentage of 67.5% (range 56%–80%) of the maximum score. Within and between-school differences in student-rated form-focused fidelity scores were much smaller than those provided by teacher, with scores ranging between 61.1% of the maximum score as rated by students in School 01 and 73.5% as rated by students in School 04. Function-focused fidelity ratings, which accumulated into an average score of 22.95 (65.7%, range 54%–76%) across schools, followed a similar pattern. Scores were lowest in School 01 (60.1%) and highest in School 04 (69.6%) but the difference was much smaller than the teacher-reported difference. The difference in scores for form and function (with higher scores for

form) is contrary to what was hypothesised at the onset of this chapter. Possible for reasons for this discrepancy are discussed later in the chapter.

Researcher-rated fidelity

Overall, both fidelity of form and function were rated much lower by the researcher than by teachers and students respectively. The average form-focused fidelity score across all observed sessions was 9.42 (36.2% of the maximum score, range 12%–73%) and functional fidelity was rated 8.47 (40.4%, range 10%–62%). Between-school differences were larger than as reported by the students but smaller than reported by the teachers, with form-focused scores ranging from 5.33 (20.5%) in School 01 to 15 (57.7%) in School 04, and function-focused scores ranging from 5.78 (27.5%) in School 01 to 12.33 (58.7%) in School 03. These findings suggest a slightly higher score for functional fidelity than for form-focused fidelity, with a somewhat smaller range in scores for functional fidelity, which is in line with the hypothesis stated at the onset of this chapter.

Inter-rater differences

Some important differences with regards to the ratings given by the different groups can be noted. First, it becomes clear from a glance at the table that the researcher rated the implementation of PBL in Studio Schools much less faithful than the students and especially teachers, in terms of form as well as function. The observation that different results may be found as a result of using different raters for rating implementation is not new, and forms a common argument for triangulation of methods when collecting implementation data (Breitenstein et al., 2010). Implications and possible explanations for these differences will be discussed in section 6.4. Second, although between-school differences on the combined activities scale are

fairly consistent across raters in terms of their direction, the difference in amplitude between the lowest and highest rating given per rater –as stated earlier– much higher for the teacher ratings than for the researcher’s, which in turn is higher than for the student ratings. Between-school differences for the ratings of the Activities A2 subscale (‘CREATE skills development’) are also inconsistent with the overall rating pattern. Third, between-school differences on the combined functions scale as rated by students form a pattern across the schools that is similar to the pattern found for the combined activities scale, but this is not the case for the teacher and researcher ratings. The inter-rater differences, in score height as well as the patterns of between-school differences per rater group, further confirm the limited inter-rater agreement that is discussed in paragraph 5.4.3, and will be further addressed in section 6.4 of this chapter.

Form-focused versus function-focused fidelity

On average across all observed sessions, the participating teachers and the researcher rater the functional fidelity of PBL implementation in Studio Schools slightly higher than fidelity to its form in terms of a percentage score (71% vs. 66% and 40.4% vs. 36.2% respectively). In contrast, ratings provided by the students were somewhat higher for form-focused than for function-focused fidelity (67.5% vs. 65.7%). This pattern was not consistent across all raters within each school, with much higher scores on the combined activities scale than on the combined function scale being reported by teachers for School 03, and by the researcher for School 04. The above suggest it is possible that fidelity of form and function are correlated, but that differences across individual sites and sessions may exist. The nature of the relationship between form- and function-focused fidelity scores will be further examined in paragraph 6.3.4 of this chapter.

6.3.3 Implementation of personal coaching in Studio Schools

Scores per rater group

The fidelity scores and percentages as defined by form and functions for personal coaching in Studio Schools, as rated by the different rater groups, are reported in Tables 6.10 and 6.11.

Coach-rated fidelity

Coaching activities were rated by coaches with an average score of 12, which represents 70.6% (range 29%–100%) of the maximum score. Similar to the teacher-submitted PBL ratings, within- and between-school differences are large, with percentage scores ranging between 56.5% as rated by coaches in School 01 to 97.1% as rated in School 04. The presence of coaching functions was rated with an average score of 19.17 by the participating coaches, corresponding to a percentage of 80.3% (range 53%–100%) . Between-school differences in reported functional fidelity range from 75.9% (as rated in School 02) to 91.1% (as rated in School 04), which represents a much smaller range than reported for the activities scale. The same pattern is found for within-school differences (see Table 6.11). These findings are in line with the hypothesis stated at the onset of this chapter.

Student-rated fidelity

Across all schools, students rated the fidelity of coaching activities with a score of 13.8, corresponding to a percentage of 81.2% (range 41%–100%) of the maximum score. Differences exist between the average ratings of students from different schools, with scores on the activities scale ranging from a score representing 73.5% of the maximum score as rated by students in School 04 versus 91.2% in School 03. Large differences also exist within schools (with sessions rated between 41 and 100%

in School 01), suggesting implementation varies within the schools. The presence of PBL functions was rated 55, corresponding to a fidelity percentage of 83.3 (range 53%–100%. Between-school differences in reported functional fidelity are smaller than those reported for form and range between 80.9% (as rated in School 01) and 90.8% (as rated in School 03). Less variability was also found for within-school differences in function-focused scores when compared to those focusing on form (the largest range of 53 to 100% reported in School 01) These findings are in line with the hypothesis formulated at the beginning of the chapter, although differences between form- and function-focused fidelity are similarly small as for the student-rated project-based learning scores.

Researcher-rated fidelity

On average, researcher-rated fidelity of form was 7.1 across observed sessions, representing a percentage of 41.8% (range 12%–82%) of the maximum score. Within and between-school differences in form-focused fidelity scores were smaller than those provided by coaches but larger than those provided by students, with scores ranging between 27.6% of the maximum score as rated for School 03 and 50.6% as rated for School 02. Function-focused fidelity ratings accumulated into an average score of 32.6 (54.3%, range 8%–75%) across schools, and followed a somewhat different pattern. Scores were lowest in School 01 (50.8%) and highest in School 04 (66.0%). The higher average score for functional fidelity than for form-focused fidelity and the smaller discrepancy in scores for the former confirm the pattern found in the coach- and student-reported scores, although larger within-school differences were found.

Inter-rater differences

Similar to the PBL ratings, the researcher's ratings of coaching activities and functions were much lower than those of the coaches and students. Students rated the form- as well as function-focused fidelity of coaching highest, as opposed to teachers for PBL. Between-school differences in fidelity ratings follow a mostly similar pattern across raters, although students rated implementation highest in School 03 as opposed to coaches and the researcher, who's average scores were highest in School 04.

Table 6.10: Personal coaching learning fidelity scores (%) as defined by form, shown per rater group

School	1	2	3	4	Total
Number of sessions observed	12	14	6	4	<u>36</u>
Avg. ratings by coaches	9.6 (56.5%)	12.6 (74.1%)	12.7 (74.7%)	16.5 (97.1%)	<u>12.0 (70.6%)</u>
(Range ratings coaches - %)	(29 – 94)	(29 – 88)	(47 – 100)	(94 – 100)	<u>(29 – 100)</u>
Avg. ratings by students	12.8 (75.3%)	14.1 (82.9%)	15.5 (91.2%)	12.5 (73.5%)	<u>13.8 (81.2%)</u>
(Ranges ratings students - %)	(41 – 100)	(53 – 100)	(76 – 100)	(59 – 88)	<u>(41 – 100)</u>
Avg. ratings by researcher	6.8 (40.0%)	8.6 (50.6%)	4.7 (27.6%)	7.3 (42.9%)	<u>7.1 (41.8%)</u>
(Range ratings researcher - %)	(24 – 76)	(12 – 82)	(12 -59)	(35 – 53)	<u>(12 – 82)</u>

Table 6.11: Personal coaching learning fidelity scores (%) as defined by function, shown per rater group

School	1	2	3	4	Total
Number of sessions observed	12	14	6	4	<u>36</u>
Avg. ratings by coaches	52.2 (79.1%)	50.1 (75.9%)	56.3 (85.3%)	60.8 (91.1%)	<u>53.0 (80.3%)</u>
(Range ratings coaches - %)	(53 – 100)	(57 – 92)	(68 – 100)	(80 – 100)	<u>(53 – 100)</u>
Avg. ratings by students	53.4 (80.9%)	53.9 (81.7%)	59.9 (90.8%)	58.0 (87.9%)	<u>55.0 (83.3%)</u>
(Range ratings students - %)	(59 – 98)	(56 – 97)	(78 – 100)	(85 – 91)	<u>(56 – 100)</u>
Avg. ratings by researcher	30.4 (50.8%)	32.6 (54.3%)	32.1 (53.5%)	39.6 (66.0%)	<u>32.6 (54.3%)</u>
(Range ratings researcher - %)	(8 – 77)	(14 – 72)	(32 – 75)	(55 – 75)	<u>(8 – 75)</u>

Form-focused versus function-focused fidelity

On average, fidelity of form and function were rated roughly in parallel, although higher scores were given on the functions scale by all raters and by the participating coaches in particular. There were several important differences between schools, with much higher function-focused fidelity scores given by some or all raters in School 01, 03 and 04, and slightly higher form-focused scores given in most schools by the students. These results suggest that in some instances functional fidelity may be higher and more consistent across different implementation contexts than fidelity as defined by form, but that in some cases functions may not always be triggered by specific activities. The nature of the relationship between form- and function-focused fidelity scores will be further examined in the next paragraph.

6.3.4 The correlation between form-focused and function-focused fidelity scores

Aggregate correlations considered per rater group

Project-based learning

Correlation coefficients for the relationship between form-focused and function-focused fidelity scores across all observed PBL sessions are presented in Table 6.12. Values ranged from .484 for the researcher ratings to .734 for the teacher ratings, representing a modest to strong correlation between scores, suggesting form-focused and function-focused fidelity for PBL in this study may be moderately to strongly related –as hypothesized– but that the correlation is not perfect. ρ was not statistically significant for the researcher ratings, although this is likely caused by the sample size of $N=12$ ratings. Coefficients for student and teacher ratings were much higher and closer to one another than the researcher rating, showing that student and teacher

scores on the PBL activities and functions scale were not only much higher than the researcher scores but also more strongly correlated.

Table 6.12: Correlation between aggregated form-focused and function-focused fidelity scores for project-based learning, shown per rater group

	Researcher (N=12)	Students (N=145)	Teacher (N=12)
Spearman's ρ (<i>p</i> -value)	.485 (.110)	.727 (.007)	.734 (.007)

Personal coaching

Table 6.13 shows the correlation coefficients for form- and function-focused fidelity scores across all observed personal coaching sessions. All coefficients were statistically significant and showed a moderate correlation between both scales, which suggests that although the scores are correlated the relationship is not very strong. This further suggests the observation made based on the PBL-coefficients that form- and function-focused fidelity do not necessarily co-occur in the study. To further assess the dynamics of this result, scores for both scales for the two studied components of the Studio Schools model were analyzed closer at the session level.

Table 6.13: Correlation between aggregated form-focused and function-focused fidelity scores personal coaching, shown per rater group

	Researcher (N=35)	Students (N=33)	Teacher (N=36)
Spearman's ρ (<i>p</i> -value)	.480 (.004)	.439 (.011)	.571 (.000)

Variation between individual sessions

Functional fidelity was measured next to fidelity of form in this study because substantial variations in the latter were expected but Studio Schools were still expected to work according to the same core principles (i.e. functions). It is therefore interesting to examine how functional and form-focused fidelity interact at the individual PBL lesson and coaching session level, because this is precisely the level at which the dynamic between both concepts becomes most relevant. Hence, scatter plots were generated for all ratings given by the researcher, the students, and the

teachers/coaches respectively to further examine how form- and function-focused fidelity ratings interact at the individual session level. To preserve a good flow, these scatter plots are not included here but instead in Appendix D.

Project-based learning

Scatter plots depicting the relationship between fidelity of form and function at the level of individual PBL lessons are presented for the researcher, students, and teachers in Appendix D1 in figures D1.1, D1.2, and D1.3 respectively. The patterns of all three are in line with the correlation coefficients presented in table 6.12, with much clearer correlations shown for the teacher and students ratings. All plots contain strong outliers (in the case of the researcher ratings, removing session 01_01 from the analysis actually increases the correlation between form- and function-focused fidelity to .592) that suggest that in some cases functions might still be fulfilled without adhering to a specific form, but also that relative fidelity to form may not always be related to functional fidelity.

Personal coaching

Appendix D2 contains the scatter plots of the relationship between form- and function-focused fidelity scores as rated by the researcher, students, and teachers in figures D2.1, D2.2, D2.3 respectively. Strongly skewed patterns towards higher student- and teacher-rated and lower researcher-rated scores confirm the large difference in fidelity ratings between these rater groups. Furthermore, as demonstrated by the much more scattered patterns in these plots and relatively greater number of outliers, greater variability exists for this component when compared to PBL with regards to the relationship between the scores on both scales. Implications

of this observation will be discussed in paragraph 6.4.2 of the discussion section of this chapter.

6.4 Discussion

This chapter reported the findings of the main component of this doctoral research project: an assessment of the fidelity with which participating Studio Schools implement project-based learning and personal coaching with respect to their form as well as functions. In this section, these findings are further discussed.

6.4.1 Current implementation fidelity in Studio Schools

Depending on the rater group (researcher, students, or teachers), form- and function-focused fidelity scores ranged from 36.2% to 67.5% and from 40.4% to 71% respectively for project-based learning, and from 41.8% to 81.2% and 54.3% to 83% respectively for personal coaching. Percentagewise, scores for functional fidelity were slightly higher than for fidelity of form.

What is the ‘right’ level of fidelity?

The fidelity percentages reported here are substantially lower than a maximum score of 100% and suggest suboptimal fidelity, but an assumption that implementation is currently failing in the participating Studio Schools based on this observation would be flawed. Researchers note that implementation fidelity never reaches 100% and seldom over 80% and positive outcomes have often been obtained with fidelity levels as ‘low’ as 60%. (Blakely et al., 1987; Botvin et al., 1990; Lipsey, 1999; Durlak & DuPre, 2008). It is clear that there is a *theoretical* imperative for Studio Schools to improve implementation fidelity as the implementation of its components is after all the method through which they seek to achieve their impact. However, without outcome data, it is difficult to determine an acceptable level of form- or function-

focused fidelity. Study limitations described in Chapter 1 prevented such data from being collected.

Within and between-school differences

Based on the submitted ratings by the teachers –the rater group that reported the largest differences in implementation between schools– between-school differences in PBL implementation differed by as much as 59.6% (form) and 40.8% (functions), and by as much as 40.6% (form) and 15.2% (functions) for personal coaching. Within schools, ratings submitted by the researcher ranges by as much as 67% (form) 70% (functions) These apparently large differences are not uncommon: between-site variability is commonly between 20% and 40% and may even be close to 90% (August et al., 2006; Durlak & DuPre, 2008). However, large differences – particularly with regards to fidelity of function– between sites working with the same program model places the extent to which the effectiveness of the model at the large scale can be assessed at risk.

Disagreement between different rater groups

The results in this chapter further confirmed the observation that the different rater groups disagreed strongly with regards to the levels of fidelity with which both assessed Studio Schools components were implemented. Even though prior research suggests that independent observations (i.e., by researchers) are more likely to be linked to outcomes than self-reported data (Hansen et al., 1991; Lillehoj et al., 2004; Resnicow et al., 1998) no outcome data were collected as part of this study hence this relationship remains unclear. It is impossible to determine which of the rater groups provided the most accurate representation of the ‘true’ rates of fidelity, and it could be argued that this is at odds with the purpose of triangulation of methods. Therefore,

any expressed preference for ratings by a specific rater group should be avoided and relative weight should instead be given to the different ratings based on the purpose (e.g., preparation for inspection, improving student experiences, staff self-reflection) for which they are being interpreted.

6.4.2 The relationship between form and functions

The essential argument underlying the distinction between form and function in thinking about implementation fidelity is the notion that functions are the main mechanisms through which specific programme activities affect outcomes, and that activities need not be standardised per se for functions to occur (Hawe et al., 2004a; Hardeman et al., 2005). In theory, this also implies that a standardised programme form may not always trigger the right functions during each session or for each participating individual. In order to empirically test these assumptions, valid joint measurement and analysis of both concepts is key. The findings discussed in paragraph 6.3.4 offer some interesting perspectives.

Higher scores for function than for form

On average, function-focused fidelity was rated slightly higher than form-focused fidelity for PBL and substantially higher (particularly by the researcher and coaches) for personal coaching. Additionally, less variation between schools was found on average for functional fidelity. These findings suggest that fidelity of function was higher and more consistent in Studio Schools than fidelity of form, which is in line with expectations for a highly flexible education model that has contextual adaptation at its roots. As noted in the discussion of the previous chapter, there is a chance that participants rated items differently on the functions scales than on the activities scales because these items often had answer options that were phrased more subjectively.

However, the researcher ratings showed a similar difference between the scales whilst using an instrument with high inter-rater reliability between objective observers. Therefore, the results seem to suggest that a distinction between implementation fidelity of form and functions can provide useful insights into the dynamics of complex interventions when these are implemented in real world settings.

The correlation between form and functions

Two core observations were made based on the analyses of the correlation between form and functions. First, the generated correlation coefficients show a moderate to strong correlation between form and function for the observed components of the Studio Schools model. This result provides some evidence for the hypothesis that form and function are not perfectly correlated, which is core to the argument for the monitoring of the latter. Moderate coefficients of around .5 also suggest that there is a chance that the micro-level logic models that were specified for PBL and coaching in Chapter 4, as well as the overall program theory of Studio Schools that was developed in Chapter 2 may not be entirely accurate and may need closer examination in further research. However, it is recognised that the outcomes of social programmes are rarely, if ever, attributable to their treatment components alone (Chen & Rossi, 1983; Pawson & Tilley, 1997), and the relationship between a programme's activities (form) and change processes (including functions) may also be affected by context (Moore et al., 2014). The second observation made based on the results presented in paragraph 6.3.4 is that the correlation between the student and teacher-rated PBL scores is much stronger than between the scores provided by the researcher, as well as for the ratings by all rater groups for personal coaching. In addition, the scatter plots generated for the personal coaching scales contained more extreme outliers (i.e. sessions in which form- and function-focused ratings were hardly correlated) than those generated for

the PBL scales. Several explanations can be thought of for this result. It could be argued that personal coaching requires an approach much more tailored to individual students than PBL. Coaching sessions aim to facilitate individual students in their learning and wellbeing and often focus on entirely different conversations about various aspects of adolescent life. Project-based learning aims to teach groups of students subjects through an independent student-directed process of enquiry. As a result, a standard set of activities may not be the right approach to fulfil individual student needs and trigger the right functions in personal coaching, whereas it may be a better option for a group-based activity such as PBL. However, the possibility that perceived social desirability or (lack of) engagement by teachers and students was more pertinent when completing the PBL questionnaires in a classroom when compared to the coaching questionnaires in a quiet and private location cannot be ruled out.

6.4.3 Limitations

Chapter 5 discussed how the measures that were used in this study carry some limitations related to their construct validity and reliability that were caused by the sampling strategy that was employed as well as decisions made in the instrument validation process. Some additional limitations should be noted that warrant caution in interpreting the results that have been presented in the previous section of the current chapter.

Studio Schools do more than project-based learning and personal coaching

For reasons explained in paragraph 4.1.1 of Chapter 4, this implementation study focused only on project-based learning and personal coaching within Studio Schools. The other core components (i.e. paid work placements and a small school

environment) as well as any additional components introduced by individual schools were not addressed. This means that limited conclusions can be drawn with regards to the implementation success of Studio Schools as a whole. It also means that any potential interaction effects of different parts of the model on the relationship between PBL and coaching activities and functions have remained unanalysed, whereas it is recognised that such analyses are very important due to the very nature of complex interventions (Craig et al., 2008; Bonell et al., 2012).

Fidelity is only one dimension of implementation

As has been highlighted before in this thesis, this study only examined fidelity of implementation, and did not assess other important dimensions such as dosage, uptake, and participant responsiveness of project-based learning and personal coaching (Durlak & DuPre, 2008; Montgomery et al., 2013). Based on the experiences of the DPhil candidate in the participating schools, these dimensions of implementation will likely vary across different Studio Schools, and may have a substantial influence on student outcomes. Initially, it was intended to measure all dimensions of implementation as to verify this claim with empirical evidence. However, early on in the project, an interesting challenge presented itself regarding the measurement of fidelity for complex interventions that require flexible implementation. Thus, the decision was made to place the focus of this doctoral project on fidelity because doing so provided an interesting opportunity to contribute to the knowledge base around the concept and the methodology used for measuring it. Because the study does not focus on the predictive capacities of implementation data with regards to outcome data, it was decided not to measure other dimensions of implementation.

Sampling issues

Whilst the sampling strategy used in the study affected the instrument validation procedures discussed in the previous chapter, it also poses some limitations for interpretation of the current results. Sampling (of schools, sessions, and participants) was done based on a rationale of convenience, which limits the extent to which the found results are generalisable to other Studio Schools, PBL and coaching sessions, and other students in the participating schools. In addition, a relatively small sample of sessions was observed. It was the intention to sample a larger number of sessions, but logistical challenges faced by the schools as well as the researcher prevented the observation of more sessions. The objectives of the study were to address an important challenge in the field of implementation and intervention research and offer insight to the Studio Schools movement with regards to their implementation practice, and the sample strategy is considered to be appropriate for these. Nevertheless, only limited generalised statements with regards to implementation in Studio Schools or the relationship between form- and function-focused fidelity can be made based on this study.

Data collection methods

Studying behaviours has been known to influence them, and collecting implementation data provides no exception (Breitenstein et al., 2010). In this study, all implementation data were collected on the basis of observations of participants, which in some cases repeatedly involved the same participants. This has likely affected participants' awareness of their implementation of and participation in sessions, which may in turn have affected the results. A second limitation comes with the fact that all implementation data were collected through questionnaires and checklists in which items were coded quantitatively on the basis of observations.

Many items constituting the form- and function-focused scales represent latent constructs such as thought patterns and attitudes, and obtaining data on these through self-report and the coding of explicitly reported qualitative passages may not have led to the most accurate representation. Because much time within the framework of this DPhil project was spent on developing the conceptual frameworks surrounding a new definition and operationalisation of the established concept of implementation fidelity, limited time could be spent on the development of innovative measures. Therefore, preference was given to developing simple but robust measures. Thus, although more sophisticated methods for data collection were considered beyond feasibility within this project, this limitation should be further addressed in the case of future assessments of form- and function-focused fidelity.

Limitations regarding the instruments used

In Chapter 5, the decision to use the same data that were used to assess implementation in Studio Schools for the purpose of analyzing the psychometric properties of the therefore used instruments was discussed. This means that conclusions drawn regarding the validity and reliability of the used instruments are not orthogonal to conclusions regarding the implementation practices that were scored with these. Time restrictions warranted this decision, but definitive conclusions on the validity of used instruments as well as implementation practices in Studio Schools would require the instruments be validated based on separate data. The additional limitations as reported in section 5.6 of the previous chapter also apply.

6.5 Conclusions

6.5.1 On implementation in Studio Schools

Comparing across schools, do the monitored components of the Studio School model currently appear to be implemented with fidelity defined by form and function?

Fidelity ratings for form (i.e., a standard set of activities) ranged from 36.2% to 67.5% for project-based learning and from 41.8% to 81.2% for personal coaching, depending on type of rater. Fidelity of function (i.e., a standard set of change mechanisms), which allows for more flexibility regarding the activities school implement, was rated slightly higher (40.4% to 71% for PBL, 54.3% to 83% for coaching). Fidelity ratings provided by students and teachers suggest that PBL and coaching are implemented with reasonable to strong fidelity when compared to ratings demonstrated to yield positive outcomes in previous research. However, much lower ratings were given by the researcher, suggesting the ratings provided by participants may have been somewhat optimistic. In addition, substantial within-school differences were found in the ratings provided by all rater groups. The current ratings suggest that work can be done by the Studio Schools Trust and the movement as a whole to further develop standard practices (at least in terms of functions), but also suggest that the model could be evaluated for its effectiveness as long as implementation and process are carefully monitored.

6.5.2 On the relationship between form-focused and function-focused fidelity

To what extent are form-focused and function-focused fidelity scores for project-based learning and personal coaching in Studio Schools correlated?

Correlation coefficients (Spearman's ρ) ranged from .485 (researcher ratings) to .734 (teacher ratings) for project-based learning, and from .480 (researcher ratings) to .571 (teacher ratings) for personal coaching. These results demonstrate a moderate yet significant (with the exception of researcher PBL ratings) correlation between form- and function-focused fidelity scores for both components of the Studio Schools model. This suggests that fidelity of form and function are correlated, but may not always be at the level of implementation of individual sessions; a hypothesis for which further evidence was found in scatter plots which demonstrated an overall pattern of correlation with several outliers. More research into the relationship between both concepts needs to be conducted in primary process evaluations of other social and health programmes in order to arrive at more definitive conclusions.

6.6 Chapter summary

Whereas the previous chapters reported conceptual frameworks and findings regarding the methodology used in this thesis, this chapter presented the findings of its empirical implementation study. Implementation fidelity of project-based learning and personal coaching was monitored in four Studio Schools over a four-month period. Although the design of the study –strongly driven by convenience sampling methodology– limits conclusions that can be drawn regarding implementation practices of Studio Schools across England, some interesting results emerged. Fidelity of function was rated somewhat higher than fidelity of form in the participating schools. Ratings provided by the researcher were much lower than those provided by teachers and students. Form-focused and function-focused fidelity scores correlated moderately to strongly, with stronger correlations found between teacher- and

student-rated scores for project-based learning. These findings suggest that fidelity of form and function are correlated to some extent (as would be expected), but may not always be at the level of individual implementation sessions. In the next chapter, more reflection will be given on the major findings of this thesis, and their implications for practice and research will be discussed. Limitations of the thesis will also be addressed.

Chapter VII: Discussion and conclusion

7.1 Introduction

This thesis intends to aid in the development and implementation of Studio Schools, and to make a valuable contribution to the scientific literature within the fields of intervention and implementation science. It focused on a study of the implementation of project-based learning and personal coaching in Studio Schools, and on improving existing as well as developing new methods for conducting implementation studies. This chapter synthesizes the lessons learned in the previous chapters, and addresses what these mean for practice and research communities that focus on effective implementation of social programmes. It starts with a summary of the main results of the thesis, after which the current state of the literature is addressed. The chapter continues with a discussion of what the thesis adds to the current knowledge base, and an examination of its strengths and limitations in doing so. It concludes by addressing the implications for future practice as well as research in implementation and evaluation of programmes.

7.2 Summary of main results

7.2.1 Specification of a programme theory for Studio Schools

Implementation measures, like other process measures, require to be based on explicit programme theory if they are to be valid. Additionally, it is important that core stakeholders in an evaluation agree with regards to what constitutes the programme under evaluation and how it is hypothesised to work before the evaluation starts. Because Studio Schools did not have an explicitly specified programme theory at the

time the DPhil project commenced, this was specified in a logic model through a Delphi-inspired process that was primarily driven by stakeholder input. Components of a logic model were nominated as items by a small panel of ten participants who play or had played an important role in the development or dissemination of the Studio School model. Participants rated these items for their relevance in a second round, after which a small majority of items were included on the basis of pre-determined criteria. The remaining items were rated as uncertain and included or excluded by the DPhil candidate on the basis of systematic criteria. This resulted in a logic model that was revised on the basis of comments received from participants, and that was subsequently approved by all participants. This logic model does not represent a 'definitive' Studio Schools logic model, and it is anticipated that it will be updated in line with developments within the Studio Schools movement. However, this study demonstrates it is feasible to involve stakeholders who do not have a background in programme evaluation in a systematic and transparent effort to specify a programme theory for evaluative purposes.

7.2.2 Reconceptualisation of implementation fidelity for complex interventions

Complex interventions such as Studio Schools comprise multiple interacting components that are constantly affected by the contexts in which these are implemented. Often, such interventions are particularly prone to local adaptations, as is the case for Studio Schools. As a result, implementation measures were developed within this thesis with the aim of providing meaningful standardised implementation data regardless of adaptation practices. Implementation fidelity was therefore reconceptualised by distinguishing between form (a specified activity or set of activities) and function (a specified set of change mechanisms resulting directly from implementation). Activities and functions were specified for each component of

project-based learning and personal coaching and were monitored in four Studio Schools for a period of four months. Analysis of the psychometric properties of the instruments used demonstrated that it is feasible to develop and validate measures that focus on form and function of implementation simultaneously. The moderate-to-strong correlations between form- and function-focused fidelity scores found in the study suggest that form and function of complex social interventions may not always correlate perfectly. Although this suggests that the hypothesis that different levels of fidelity of implementation of standardised activities may not correspond directly to differences in functional fidelity may be accurate, alternative explanations of this finding (e.g., imperfect programme theory, measurement error) should also be considered.

7.2.3 Assessment of implementation fidelity in Studio Schools

Implementation fidelity of project-based learning and personal coaches was monitored through paper-based quantitative checklists completed by the DPhil candidate, participating staff, and students. Form- and function-focused fidelity scores ranged from 36.2% to 67.5% and from 40.4% to 71% respectively for PBL, and from 41.8% to 81.2% and 54.3% to 83% respectively for coaching. Scores were slightly higher for function than for form. Between-school differences in implementation fidelity were very large, which is not uncommon for social interventions. Although an acceptable level of fidelity can only be determined by assessing the relationship between implementation and outcome data, the substantial heterogeneity between the implementation practices –as defined by form as well as function– found within the participating schools warrants careful consideration of implementation in future effectiveness studies. Large differences also exist between the ratings supplied by

different rater groups, confirming the value of triangulation within implementation monitoring, and suggesting that the used instruments can be improved.

7.3 What we already knew about the addressed topics

7.3.1 Programme theory development

The need for social intervention programmes and their evaluations to be based on explicit, specific, and evaluable programme theories is widely acknowledged in the literature (Chen & Rossi, 1983; Rogers, 2007; Weiss, 1997). Although the core argument for this lies in the improved explanatory power of theory-driven evaluations (how, why, and for whom do programmes work under what circumstances), it brings along additional benefits. Explicit programme theories may expose weaknesses in programme design and can aid in the recognition and prevention of potential adverse effects (Bonell et al., 2015; Moore et al., 2014). Likely a result of their basis in a grounded understanding of the processes and contexts that affect human development and behaviour, there is also increasing evidence that theory-based interventions are more effective than those not based on explicit theory (Glanz & Bishop 2010; Noar et al., 2008). Finally, clear programme theories facilitate programme implementation through clarifying the essential resources needed for success, and may resolve conflicts in perceptions of core stakeholders regarding intervention models (Cooksy et al., 2001; Moore et al., 2014). Nevertheless, many evaluations still fail to specify and test programme theories explicitly, resulting in a loss of opportunity to better understand, evaluate, improve, and implement them (Bonell et al., 2012; Gardner et al., 2010; Trifiletti et al., 2005). This may be particularly the case when programmes are developed by non-evaluators.

7.3.2 The fidelity versus adaptation debate

As an essential part of the intervention process and the primary driver of programme effects, implementation lies at the heart of theory-driven process evaluations. Although implementation comprises multiple aspects (including for example dosage, quality, uptake, and participant responsiveness), fidelity is the focus of much of the debate in the implementation literature. As it has been demonstrated that implementation fidelity is positively related to intervention outcomes, the need to implement programmes as developed is widely recognised (Breitenstein et al., 2010; Durlak & DuPre, 2008; Mihalic, 2004). To monitor implementation fidelity in evaluations and to improve the implementation and dissemination of effective interventions, their components require to be standardised to some extent. Although this standardisation is welcomed by some as a means to ensure effective implementation (Webster-Stratton, 2004), others argue it leaves insufficient room for the tailoring or adaptation of services to different contexts, thus potentially compromising the effectiveness of an intervention (Kumpfer et al., 2002; Castro et al., 2004). In recent years, the ‘fidelity versus adaptation debate’ has become more nuanced, with authors recognising these concepts need not be mutually exclusive. Shifting the focus of intervention standardisation for complex interventions to the change mechanisms (functions) in a programme theory versus the specific form of a programme is one way of allowing for contextual adaptation whilst maintaining fidelity to a programme theory (Hardeman et al, 2005; Hawe et al., 2004a). To accomplish this, new thinking about what constitutes the implementation of social interventions as well as explicit reporting of the programme theorization process is paramount.

7.3.3 Research into Studio Schools

Studio Schools are a new model of secondary education in England that focuses on teaching students core GCSE subjects and transferable non-cognitive skills through an applied and contextualised pedagogy (Studio Schools Trust, 2012a). The schools do not work with a specific set of admission criteria except for the suitability of their unique approach to a student's style of learning. The model aims to target youth disengagement with secondary education as well as the poor employability of secondary school leavers, and there are currently 40 schools in operation across England. Although the schools combine instruction techniques and additional components that have been 'tried and tested', the programme theory underpinning the model is not described anywhere in the academic literature, nor has any research been conducted on its implementation or effectiveness in Studio Schools prior to this DPhil project.

7.4 What this thesis adds

7.4.1 Theory-driven evaluation, implementation fidelity, and adaptation

Evaluators commonly need to specify their theories after these are designed but prior to their evaluation (Connell & Kubisch, 1999). Examples of such efforts are reported in the literature (e.g., Christie & Alkin, 2003; Sullivan et al., 2002), but the methodological processes for elucidating a programme theory are rarely specified a priori and reported explicitly. Hence, it is often unclear how an end product (e.g., logic model) resulted from researchers' efforts and how these may have affected its validity. Additionally, poor reporting limits the development of best practices for elucidating implicit assumptions regarding programme theories, which forms a core challenge at the onset of evaluation efforts that heavily rely on stakeholder

participation, such as for example pragmatic trials. This thesis has demonstrated that a systematic consensus process can be employed to involve stakeholders without academic backgrounds in programme theory specification. It involved a thoroughly participatory approach towards identifying the core elements of a programme's logic model, although it was more researcher-driven in the identification of the causal relationships between these elements. Furthermore, it offers a detailed reconstruction of the process (in the form of an online supplement that is also included in Appendix A) and critically reflects on it, so that it can be replicated and improved in future efforts (Van Urk et al., 2015).

Defining and operationalising functions

In order to improve our understanding of how and why social programmes work or do not work, authors have repeatedly argued for better theorization of interventions (Chen & Rossi, 1983; Weiss, 1995; Rogers, 2007). To allow for the implementation of complex interventions to respond to the complexity of the settings in which they are implemented, authors have also called for a new way of thinking about what constitutes the essence of these interventions and how they should therefore be standardised (Hawe et al., 2004a). Core to these arguments are functions, or mechanisms of change, that result from implementation and that produce outcomes in participants. However, many different understandings of mechanisms exist, and the literature is not clear on how to link them to intervention strategies or activities implemented in empirical studies (Lacouture et al., 2015). This thesis adds to the literature on theory-driven evaluation as well as implementation science by conceptualising, operationalising, and empirically testing a new approach towards improved theorization of the intervention implementation process. It builds further on previous efforts to disentangle intervention techniques from the change mechanisms

that they aim to facilitate and that are rooted in programme theory (Gardner et al., 2010; Hardeman et al., 2005; Michie et al., 2008a; 2008b), but offers more implications for implementation research by clearly explaining and demonstrating how our thinking about implementation fidelity is affected by a distinction between form and function. It has defined programme components as comprising activities and functions, offered clear distinctions between the three concepts, and explained that functions differ from other mechanisms of change in that they form the unique and immediate contribution of activities to a programme theory. This allowed for the operationalisation of activities and functions of project-based learning and personal coaching in Studio Schools into conceptually separate constructs that were monitored simultaneously. An empirical assessment of the psychometric properties of the instruments developed based on this operationalisation showed that although important limitations apply, it is possible to treat activities and functions as empirically separate constructs that nevertheless somewhat correlate. The instrument assessment process also demonstrated that researchers can code the implementation of activities and the presence of functions reliably for complex interventions. Through this process, the thesis has demonstrated the possibility of focusing implementation monitoring on function next to form, resulting in the latter needing less standardisation (i.e. ‘packaging’) than is sometimes assumed in the field (e.g., Mihalic, 2004), without forfeiting the detailed rigour of process evaluations and implementation studies. It has therefore helped make empirical a debate that was previously mostly theoretical, and has offered new options to alleviate the tension between implementation fidelity and adaptation, as well as improving the systematicity of programme theory improvement based on the empirical findings of primary studies (see paragraph 3.5.2 of Chapter III). Whether developers and

evaluators of complex interventions should strive to optimise fidelity of function as opposed to form is currently unclear and can only be verified by assessing empirically which type of fidelity is a better predictor of programme effectiveness in future studies.

7.4.2 Studio Schools

This thesis also represents the first scientific research conducted on Studio Schools. A clear programme theory was developed through a rigorous, well-documented process, where information regarding the Studio School programme theory previously only existed implicitly and scattered in information documents (Studio Schools Trust 2012a; 2013b; Van Urk et al., 2015). Although the model as defined in this thesis will likely be further developed and revised, the schools and their evaluation partners can use the work produced as a starting point for evaluation design (Eldridge et al, 2005). Furthermore, insight is provided by the thesis in the implementation of the model within the schools, which can lead to improvements made to current practice but also indicates the attention that needs to be given to implementation heterogeneity in future effectiveness trials on Studio Schools. Through these efforts, this thesis has contributed substantially to the initial knowledge based surrounding Studio Schools and has improved the evaluability of the model, so that performance can be improved, the planning of outcome evaluations can be done more efficiently, and the results of such efforts will gain more explanatory power (Ogilvie et al., 2011; Wholey 1996).

7.5 Strengths and limitations

Innovative use of methodology

This thesis offers a new definition of intervention programme components, and a new conceptualisation of implementation fidelity for complex interventions based on this

definition. It is acknowledged by the DPhil candidate that the contribution of this thesis lies not so much in developing concepts that are entirely new: the concepts of intervention components, functions, activities or techniques, and mechanisms of change are all established concepts in the evaluation literature that have been derived from previous work (Chen & Rossi, 1983; Hawe et al., 2004a; Pawson & Tilley, 1997; Michie et al., 2008; Webster-Stratton, 2004; Weiss, 1995). However, this thesis has analysed and presented the conceptual distinctions and interdependencies between these concepts more explicit than has been done before, so that they could be operationalised and utilised to allow for more flexibility in implementation practices as part of process evaluations of complex interventions. Furthermore, it built further on the conceptualisation process by a subsequent operationalisation and empirical analysis of the validity and reliability of instruments that were based on it. This process was carried out systematically and reported explicitly in terms of its underlying assumptions, methodological decisions, and limitations. Thus, it can be replicated and improved based on lessons learned from this effort, increasing the possibilities for a much needed accumulation of empirical evidence regarding the validity and utility of improved theorization of interventions. Such a systematic and explicit discussion of concepts, operationalisation, and implications for empirical use is not always offered when new approaches towards (aspects of) programme evaluation are proposed, including in previous work on the relationships between functions and activities within programme theories (Hawe et al., 2004a; Michie et al., 2008).

Limitations

The previous paragraph is not to say that this thesis is without limitations. First, parts of the nature of the relationship between activities and functions were left unexplored

due to a limited research design. For example, although correlations between activities and functions in PBL and coaching implementation in Studio Schools were demonstrated, no claims can be made at this point regarding the causation of any of these concepts by another. Additionally, because no outcome data were included in the analyses, no data are available to inform a discussion on the hypothesised pathways from activities to outcomes via functions. Therefore, the strengths of this thesis are in its demonstrative value. Second, some decisions were made during the planning and carrying out of the study that have likely affected its results. The most notable of these was the decision not to include a second Delphi round when developing the Studio Schools logic model (see Chapter II), a sampling strategy for selecting participating schools, participants, and sessions for implementation monitoring that was based on convenience (see Chapters V and VI), and decisions made during the data analysis process based on the aim to optimally assess the conceptual hypotheses in this thesis instead of developing optimal measures for Studio Schools (see Chapter V). The possible influences of these decisions on the results and conclusions of the study are addressed in detail in Chapters II, IV, V, and VI, and can be summarized as limiting the extent to which definitive conclusions can be drawn regarding the validity of function-focused monitoring of implementation fidelity, as well as the current state of implementation of project-based learning and personal coaching in Studio Schools. It is acknowledged here that, although offering an interesting and promising approach towards implementation fidelity, this thesis does not pretend to provide a magic bullet for increasing the flexibility of standardisation of the content of complex interventions. It is currently not yet clear how useful this approach is towards theory-driven evaluation and fidelity monitoring, and more advanced research designs will be needed to remove this lack of clarity.

Stakeholder consultation

An additional strength of this thesis lies in the extensive consultation of programme stakeholders throughout the research process. Because Studio Schools represent a new education model that is being implemented before any programme theory or methods for evaluation were designed, it was necessary to involve their stakeholders in the development of these. Developers, implementers, teachers, and students of Studio Schools were involved, and the methods used to obtain and use their inputs are reported explicitly. Various sources in the literature stress the importance of collaborating with programme stakeholders during evaluations (Julian, 1997; Krieger et al., 2002; Oosthuizen & Louw, 2013; Sullivan et al., 2002). Involving stakeholders in evaluations is paramount to ensure that research is rooted in a shared understanding of programmes between researchers and the implementers and users of these services. It is also essential in making sure that the results of evaluations are of benefit to those that can use them to improve services. This study employed iterative, systematic, and explicitly reported methods for collaborating with Studio Schools' stakeholders in improving its validity and utility. This too, however, carries some limitations. As discussed more extensively in Chapter II, the iterative consultation process that was used to develop a logic model for Studio Schools caused some delay in the theory development process, and did not necessarily make it more rigorous. The relatively tedious exercise warranted allowed less time than was originally planned for the remainder of the study, including data collection. Repeated revision of the used measures with the help of the Studio Schools Trust, teachers and students added somewhat onto this, although this was considered necessary to develop instruments with strong content validity. These delaying factors introduced some of the limitations discussed earlier in this section as well as in the previous chapters, and therefore

affected the extent to which definite conclusions that can be drawn based on the study.

7.6 Implications for research

Above all, this thesis represents a work in progress. It is demonstrative in purpose, and seeks to open new doors towards thinking about implementation fidelity and process evaluations. Whether or not the approach developed as part of this doctoral project –that of a function-focused approach towards monitoring implementation fidelity– will be of sustained benefit to evaluation science will depend on its further development, improvement, and ultimately its utility in future studies.

7.6.1 Implications for process evaluation and the monitoring of implementation

Several implications of this thesis for implementation monitoring and process evaluation can be discussed. Primarily, there is a need for more empirical enquiry (i.e. analysis of data collected in the field with the help of operationalised concepts) into what constitutes the components of complex social interventions and how the relationships between the implementation of activities (or intervention techniques), functions, subsequent mechanisms of change, and outcomes can best be modelled within process evaluations. Current work on what constitutes interventions and how their components could be standardised in order to develop evidence-based practices around promising models that take contextual influences into account is mostly theoretical (e.g., Hawe et al., 2004a; Lacouture et al., 2015; Rogers, 2008). The development of valid approaches towards the standardisation of programme theories rather than activities ('packages') is dependent on the accumulation of empirical evidence for and against the validity and utility of promising approaches (Dalkin et

al., 2015). This work need not be very complicated: relatively simple exercises such as completed in this thesis may provide proof of concept for interesting ideas, after which larger studies can provide more insight into their utility and validity.

Future research could build on the specific approach that was developed within this thesis in several ways. First, improved research designs could be used to better assess the three core assumptions that support a function-focused understanding of implementation fidelity as originally presented in Box 3.1 (repeated below).

1. Occurring functions are the result of implementation of an activity or set of activities, and the extent to which the presence of specified functions correlates to the implementation of the activity/activities needs to be demonstrated empirically
2. Provided that all functions of all implemented activities are specified appropriately in a valid theory of change, the relationship between implemented activities and outcomes is mediated in its entirety by the functions of the activities
3. The extent to which specified functions are related to specified outputs and outcomes corresponds to the validity of programme theory, and not to the specific form of an implemented set of programme activities

Box 3.1 (repeated): Three underlying assumptions of a function-focused approach towards defining and monitoring implementation fidelity

To properly assess all three assumptions, a multisite randomised factorial design (in which specific activities would be implemented in some arms but not others) would be required to test for any causal effects of a specific form (i.e. different combinations of specific activities) on functions (assumption one) (Shadish et al., 2002). An appropriate sample size would be required to test assumptions two and three through mediation analyses (for example a pathway analysis in the form of a SEM) that would lie at the heart of process evaluations (Preacher et al., 2007). Naturally, outcome data should also be included in such a study. It is recognised here as a limitation of this approach that such research designs are amongst the most resource intensive and may not always be feasible. Second, repetition of factor analyses such as the one presented

in Chapter V in future studies could help confirm or contest the finding that activities and functions can load as separate, yet correlated constructs, and would improve our understanding of how these relate to one another. Other methods commonly used to assess the psychometric properties of instruments (e.g., principal component analysis) might also be employed for this purpose. Third, researchers could explore the use of instruments other than questionnaires to collect quantitative function-focused implementation data in order to further decrease/rule out risks of monomethod bias and other method problems (Bank et al., 1990; Cook & Campbell, 1979; Dishion et al., 1999). Explicit self-reports on implicit constructs may not provide the most valid data on participants' attitudes, beliefs, behaviours, and other constructs that may often classify as functions, and other measures should be explored. Fourth, a major limitation of the current study was the absence of qualitative findings to further explain its quantitative results, and future research should focus on collecting qualitative data to add further richness and detail to these results as to better understand how and why fidelity of form may or may not produce its relevant functions and how contextual factors may affect this relationship (Basch et al., 1985; Oakley et al., 2006). Such efforts could focus on a deeper understanding of how functions occur in the minds of programme staff and participants, and on the relationship between a standard set of functions and a variety of different activities or intervention techniques. Lastly, efforts similar to the study reported in this thesis but from other fields than education (e.g., public health, psychology, or management sciences) would be welcomed in order to explore the applicability of concepts that were developed in the context of the current project in other domains that can benefit from innovations in evaluation methodology.

Another direction for future research that is implied by this thesis is to investigate further which types of interventions might benefit from an approach towards implementation monitoring that focuses on functions, versus types of interventions for which a traditional (form-focused) approach might be more appropriate. Several dimensions of programmes may affect whether an evaluation of their implementation should focus primarily on their functions or their form, including their complexity, the contexts in which they are implemented, and the diversity of participants they intend to reach. Only empirical research that includes analyses on the predictive power of form- versus function-focused fidelity for programme effectiveness can determine what type of fidelity to focus on for a given intervention. A key role in this is recognised for qualitative research with programme staff and participants, in which their understandings and beliefs of activities, functions, and their relevance for the delivery of and participation in intervention programmes are explored.

7.6.2 Implications for stakeholder involvement and participatory research

In order to use limited resources to conduct rigorous and valid research that produces results that optimally inform practice, researchers will need to integrate the need to involve stakeholders in programme evaluation with the need to efficiently carry out research of scientific standard. Explicit reporting of selected approaches to undertake such an effort is essential to work towards best practices. Chapter II argued that, in line with general suboptimal reporting in trials of social- and psychological interventions (Grant et al., 2013), reporting of efforts to engage stakeholders in research is poor, and improvements are required.

Although the Delphi-inspired process for programme theory elucidation that was discussed in Chapter II forms one example of an approach that could be repeated and

improved as a result of its explicit planning and reporting, many alternative methods and aims of these methods can be explored. Advocates for participatory research have made strong cases for the involvement of (often marginalised) communities in formulating research questions, selecting appropriate evaluation methods, and interpreting research results (for example: Diaz & Simmons, 1999; Green & Mercer, 2001). It is currently unclear if systematic consultation processes (e.g., Delphi, Nominal Group Technique) form the most optimal method to engage communities in research focusing on their circumstances or the improvement of these. Yet, using such methods for this purpose deserves more exploration as their egalitarian approach towards participation aligns with the “commitment to the sharing of power with the people with and for whom researchers work” that lies at the heart of participatory research (Cornwall & Jewkes, 1995). The exact method of choice for facilitating stakeholder involvement in evaluation efforts will depend on the purpose of the process and the relationships between participants and between them and the researcher. However, a commitment to systematicity and transparency when conducting and reporting such efforts is warranted.

7.6.3 Implications for future research on Studio Schools

This thesis provides Studio Schools with an initial knowledge base around their programme theory as well as the implementation of some core curricular components in a select sample of schools. In addition, it has developed a set of measures specifically developed for these components and for Studio Schools that were developed and refined through a rigorous process. Future research on this education model should focus on effectiveness, but should also seek to develop to increase understanding of how different parts of the model interact to produce outcomes and

what the defining features (e.g., activities, functions, change mechanisms) of the Studio School model are after some years of implementation. Specifically, a number of directions for research in the schools can be recommended. First, the programme theory that was specified in the form of a logic model in Chapter II could be updated based on implementation experiences since it was developed, as well as qualitative insights obtained from teachers and students. Second, a pilot effectiveness trial should be designed and conducted to assess the efficacy of the method compared to alternatives. Such a trial could also be employed to explore the best feasible design for evaluating the effectiveness of Studio Schools. Where more students apply for entry than can be admitted in multiple Studio Schools, a randomised design might be employed, but the recent closing of two schools as a result of shortage of students suggests this may be difficult (McKinney, 2015). Alternative (quasi-experimental) designs should therefore also be considered. Finally, future research on Studio Schools should retain a focus on process and implementation. The expected heterogeneity of implementation between schools that was confirmed in this study –in terms of form as well as function– warrants that care is taken to understand how the Studio School model affects students depending on different patterns of implementation. Widely ranging (from ‘Outstanding’ to ‘Inadequate’) inspection results further confirm this need (OFSTED, 2015a; 2015b; 2015c; 2015d).

7.7 Implications for practice in Studio Schools

Paragraph 6.5.1 of the previous chapter concluded that implementation of project-based learning and personal coaching is being implemented with reasonable fidelity (both in terms of form and function), based on ratings given by teachers and students, when compared to ratings of fidelity found in studies of other social interventions.

Ratings given by the researcher were much less high, and substantial differences existed between the schools. There is no definition in the literature of an optimal level of fidelity, and because no outcomes were included in the implementation analyses conducted in this thesis, no recommendations regarding the right level of fidelity can be given per se. Some implications nevertheless apply.

7.7.1 Ensure consistency of the Studio School approach

Currently, support given to Studio Schools by the Studio Schools Trust, other schools, or other bodies to guide implementation, is limited to voluntary training sessions and resource sharing. The current study revealed large discrepancies in implementation fidelity (compositionally as well as functionally) of PBL and coaching between the participating schools. As has been argued throughout this thesis, the need for some standardisation in practices across sites is required for a model to be recognisable and to ensure effectiveness. To ensure consistent implementation of the model and a quality education for all students enrolled in Studio Schools in England, much further thought will need to be given to what constitutes the essence of the Studio School model and its individual components, and the implications this should have for implementation of the model in individual schools. Although Chapters II, III, and IV of this thesis represent a systematic effort to define the model's programme theory and measurement instruments for evaluation purposes, the Studio Schools movement is developing on an ongoing basis. The model is now being implemented for several years, new schools open each academic year and new stakeholders join continuously. Thus, updated implementation guidance that incorporates actual implementation experiences (as opposed to mere social scientific theory) and that is based on consensus amongst the wider Studio Schools community will be essential to ensure a universal model that has consistent pedagogic quality. Such guidance could be based

on a combination of activities and functions, depending on the flexibility in implementation that is deemed necessary for specific components of the model. It should also include strategies for linking specific activities to functions and how activities could be adapted whilst (in theory) leaving functions intact. As the central body that links all Studio Schools and their civil partners, the Studio Schools should play a key role in developing and disseminating such guidance. Consistent implementation guidance should go hand in hand with consistent evidence-informed quality improvement efforts, which will require metrics such as those developed in this thesis for all active components of Studio Schools. The Studio Schools Trust is currently developing a universal metric for monitoring the development of students' CREATE skills in the schools, and more rigorously developed metrics should follow. To facilitate adequate inspection of Studio Schools by educational authorities that does justice to their unique pedagogy, communication of such guidelines to and perhaps involvement of OFSTED is recommendable.

7.7.2 Build appropriate levels of organizational capacity

More attention will need to be given to the organizational capacity of individual Studio Schools as well as the delivery and support systems (comprising the schools, the SST, funders, and other collaborating partners) as a whole. As a community-centred model (i.e. one that begins with the need of a community and uses scientific information to produce an effective model), the capacity of schools and the Studio Schools system to effectively deliver their approach is key to improve their functioning (Bond et al., 2001; Durlak & DuPre, 2008; Wandersman et al., 2008). Two types of capacities are specifically required: the capacity required to deliver the Studio School model in terms of skills and motivation and a general capacity for effective organizational structure and functioning that promotes the former (Livet &

Wandersman, 2005). The Interactive Systems Framework that was developed by Wandersman and colleagues and that is commonly used in the implementation literature explains that to effectively deliver prevention models requires not only a focus on the capacity of prevention deliverers (i.e. teachers, principals of Studio Schools) but also on the prevention support system (in the form of funders, overarching academies, the Studio Schools Trust and collaborating businesses). Anecdotes obtained from conversations with principals and teachers in the participating Studio Schools suggest that capacity in the form of specialised teachers (e.g., who are comfortable with coaching young people and with teaching curricula through projects) and in some cases sheer manpower may not be sufficient in some Studio Schools. The closing of two schools resulting from a lack of funding due to under subscription further suggests that not all required resources to work effectively are in place in every school, whilst the Studio Schools Trust is also limited in its practices by scarce funding. The results of this research demonstrate that Studio Schools have potential, but that individual schools require different funding arrangements that recognise and provide for their small scale (i.e. high teacher-to-pupil ratios) and personal learning environment (i.e. additional staff specialised in personal coaching) to be sustainable. This implies that the common 'per pupil' funding arrangement that is in place for British secondary schools may not be appropriate for Studio Schools. Now that Studio Schools are operational for some years, an adequate needs assessment in terms of their delivery system as well as their support system as well as a strategy for meeting these is recommended. Such a strategy should at least include specialised training and ongoing support for all Studio Schools staff related to the pedagogic approaches used in all Studio Schools (i.e., PBL, personal coaching, work placements), as well as ongoing (internal) monitoring

of teaching practices to maintain fidelity of implementation and for quality improvement purposes. Currently, the SST only offers centralised training sessions on a voluntary basis, which are often attended by representatives of local schools rather than their entire staff. This practice brings along a risk of adapting standard practices not in order to facilitate a good match with local contexts, but rather due to selective or erroneous repetition of training sessions in individual schools. To ensure a sustainable quality package of training and ongoing support, the Studio Schools Trust will require a substantial increase in funding so that additional training staff can be hired and supported.

7.8 Closing remarks

Evidence-based practice in the public domain was introduced because of the imperative of services to do more good than harm for their users, and to ensure that public funds are used as effectively as possible. Without successful implementation of well-evaluated initiatives, whether developed by communities or researchers, the success of evidence-based interventions and policies cannot be assumed. It is argued in this thesis that some standardisation of intervention programmes is required to facilitate their evaluation and replication. However, an individualised world demands that social services be flexible and tailored to some extent to the needs of individuals and their specific communities and settings. Without these features forming an integral part of the design, evaluation, and implementation of these programmes, there is a risk that promising models may not be transferable. As a result, evidence generated in specific contexts is of little use in others, and much of its potential is lost. This doctoral thesis stresses the importance of empirical work on the nature of programmes and their components, as well as on the relationships between the

elements that constitute these and their outcomes. It does not pretend to offer a magic bullet for improving intervention design and implementation, but may form a useful brick in a movement within evaluation and implementation science that aims to improve the utility and thus use of research evidence by policy makers. It is the hope of the doctoral candidate that this work may inform improved efforts to develop and evaluate intervention models that are informed by sound scientific evidence and can be successfully tailored to and implemented in the different contexts in which they are needed.

References

- Academies Act 2010 (c. 32). London: HMSO.
- Alexander, P.A. & Murphy, P.K. (1999). Nurturing the seeds of transfer: a domain-specific perspective. *International Journal Educational Research* 31: 561-76.
- Amadeo, C. & Marshall, P. (2013). Low educational attainment in England: an audit. In: Marshall, P. (ed). *The Tail: How England's schools fail one child in five – and what can be done*. London: Profile Books.
- Arthur, M.W. & Blitz, C. (2000). Bridging the gap between science and practice in drug abuse prevention through needs assessment and strategic community planning. *Journal of Community Psychology* 28: 241-55.
- Arum, R. & Shavit, Y. (1995). Secondary Vocational Education and the Transition from School to Work. *Sociology of Education* 68(3): 187-204.
- Audrey, S., Holliday, J., Parry-Langdon, N., & Campbell, R. (2006). Meeting the challenges of implementing process evaluation within randomized controlled trials: the example of ASSIST (A Stop Smoking in Schools Trial). *Health education research* 21(3): 366-77.
- August, G. J., Bloomquist, M. L., Lee, S. S., Realmuto, G. M., & Hektner, J. M. (2006). Can evidence-based prevention programs be sustained in community practice settings? The Early Risers; advanced-stage effectiveness trial. *Prevention Science* 7: 151– 165.
- Avery, A.J., Dex, G.M., Mulvaney, C., Serumaga, B., Spencer, R., Lester, H.E. & Campbell, S.M. (2011). Development of prescribing-safety indicators for GPs using the RAND Appropriateness Method. *British Journal of General Practice*: e.526-36.

- Backer, T.E. (2001). *Finding the balance –Program fidelity and adaptation in substance abuse prevention: a state-of-the art review*. Center for Substance Abuse Prevention, Rockville, MD.
- Baglin, J. (2014). Improving your exploratory factor analysis for ordinal data: a demonstration using FACTOR. *Practical Assessment, Research & Evaluation* 19(5): 2.
- Bailey, T. & Hughes, K. (1999). *Employers Involvement in Work-Based Learning Programs*. Berkeley, CA: National Center for Research in Vocational Education.
- Bank, L., Dishion, T., Skinner, M., & Patterson, G. R. (1990). Method variance in structural equation modelling: Living with" glop.". *Depression and aggression in family interaction*: 247-279.
- Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of personality and social psychology* 51(6): 1173.
- Barrows, H.S. (1996). Problem-Based Learning in Medicine and Beyond: A Brief Overview. *New Directions for Teaching and Learning* 68: 3-12.
- Barrows, H. S. (2002). Is it Truly Possible to Have Such a Thing as dPBL? *Distance Education* 23(1): 119-122.
- Basch, C. E., Sliepcevich, E. M., Gold, R. S., Duncan, D. F., & Kolbe, L. J. (1985). Avoiding type III errors in health education program evaluations: a case study. *Health Education & Behavior* 12(3): 315-331.
- Battistich, V., Schaps, E., Watson, M., Solomon, D., & Lewis, C. (2000). Effects of the Child Development Project on students' drug use and other problem behaviors. *Journal of Primary Prevention* 21(1): 75-99.

- Beinart, S., Anderson, B., Lee, S. & Utting, D. (2002) *Youth at Risk? A national survey of risk factors, protective factors and problem behaviour among young people in England, Scotland and Wales*. London: Communities that Care.
- Bellg, A.J., Borelli, B., Resnick, B., Hecht, J., Minicucci, D.S., Ory, M. *et al.* (2004). Enhancing treatment fidelity in health behavior change studies: Best practices and recommendations from the NIH behavior change consortium. *Health Psychology* 23: 443-451.
- Bereiter, C. (1997). *Situated cognition and how to overcome it*. In D. Kirshner & J.A. Whitson (Eds.), *Situated cognition: Social, semiotic, and psychological perspectives* (pp. 281-300). Hillsdale, NJ: Lawrence Erlbaum.
- Bernal, G. & Saez-Santiago, E. (2006). Culturally Centered Psychosocial Interventions. *Journal of Community Psychology* 34(2): 121-32.
- Berwick, D.M. (2003). Disseminating Innovations in Health Care. *JAMA* 289(15): 1969-75.
- Blakely, C. H., Mayer, J. P., Gottschalk, R. G., Schmitt, N., Davidson, W. S., Roitman, D. B., et al. (1987). The fidelity- adaptation debate: Implications for the implementation of public sector social programs. *American Journal of Community Psychology* 15: 253–268.
- Bloom, H.S., Thompson, S.L. & Unterman, R. (2010). *Transforming the High School Experience. How New York City's New Small Schools Are Boosting Student Achievement and Graduation*. New York, NY: MDRC.
- Bolton, P. (2013). *Grammar school statistics – Commons Library Standard Note* SN01398. Accessed 3 December 2013 at <http://www.parliament.uk/briefing-papers/SN01398>

- Bond, L., Glover, S., Godfrey, C., Butler, H. & Patton, G.C. (2001). Building Capacity for System-Level Change in Schools: Lessons from the Gatehouse Project. *Health Education & Behavior* 28: 368-83.
- Bond, L., Butler, H., Thomas, L., Carlin, J., Glover, S., Bowes, G., & Patton, G. (2007). Social and School Connectedness in Early Secondary School as Predictors of Late Teenage Substance Use, Mental Health, and Academic Outcomes. *Journal of Adolescent Health* 40: 357e9-357e18.
- Bonell, C., Fletcher, A., Morton, M., Lorenc, T. & Moore, L. (2012). Realist randomised controlled trials: A new approach to evaluating complex public health interventions. *Social Science & Medicine* 75: 2299-2306.
- Bonell, C., Fletcher A., Fitzgerald-Yau, N., Hale, D., Allen, E., Elbourne, D., Jones, R., Bond, L., Wiggins, M., Miners, A., Legood, R., Scott, S., Christie, D. & Viner, R. (2015). Initiating change locally in bullying and aggression through the school environment (INCLUSIVE): A pilot randomised controlled trial. *Health Technology Assessment* 19(53).
- Bonell, C., Jamal, F., Melendez-Torres, G.J., & Cummins, S. (2015). 'Dark logic': theorizing the harmful consequences of public health interventions. *Journal of Epidemiology & Community Health* 69: 95-8.
- Bonell, C., Oakley, A., Hargreaves, J., Strange, V. & Rees, R. (2006). Assessment of generalisability in trials of health interventions: suggested framework and systematic review. *BMJ* 333: 346-9.
- Bonell, C, Allen, E., Strange, V., Copas, A., Oakley, A., Stephenson, J. & Johnson, A. (2005). The effect of dislike of school on risk of teenage pregnancy: testing of hypotheses using longitudinal data from a randomised trial of sex education. *Journal of Epidemiol Community Health* 59: 223-30.

- Bonell, C.P., Strange, V.J., Stephenson, J.M. et al. (2003). The effect of various dimensions of social exclusion on young people's risk of teenage pregnancy: development of hypotheses from analysis of baseline data arising from a randomized trial of sex education. *Journal of Epidemiol Community Health* 57: 871–6.
- Borrelli, B., Sepinwall, D., Ernst, D., Bellg, A.J., Czajkowski, S., Breger, R., DeFrancesco, C., Levesque, Sharp, D.L. & Ogedegbe, G. (2005). A New Tool to Assess Treatment Fidelity and Evaluation of Treatment Fidelity Across 10 Years of Health Behavior Research. *Journal of Consulting and Clinical Psychology* 73(5): 852-60.
- Botvin, G.J., Baker, E., Filazolla, A.D. & Botvin, E.M. (1990). A Cognitive-Behavioural Approach to Substance Use Prevention: One-Year Follow-up. *Addictive Behaviors* 15: 47-63.
- Botvin, G.J. & Griffin, K.W. (2004). Life Skills Training: Empirical Findings and Future Directions. *Journal of Primary Prevention* 25(2): 211-32.
- Bowen, N.K. & Guo, S. (2011). *Structural Equation Modelling*. Oxford: Oxford University Press.
- Bowles, S. & Gintis, H. (1976). *Schooling in Capitalist America: Educational Reform and the Contradictions of Economic Life*. New York: Basic Books.
- Bowles, S. & Gintis, H. (2002). Schooling in Capitalist America Revisited. *Sociology of Education* 75(1): 1-18.
- Boyce, W., Torsheim, T., Currie, C., & Zambon, A. (2006). The family affluence scale as a measure of national wealth: validation of an adolescent self-report measure. *Social Indicators Research*, 78, 473-87.

- Braconier, H. (2012). *Reforming Education in England. Economics Department Working Paper No. 939*. Organisation for Economic Co-operation and Development.
- Breitenstein, S.M., Gross, D., Garvey, C.A., Hill, C., Fogg, L. & Resnick, B. (2010). Implementation Fidelity in Community-Based Interventions. *Research in Nursing and Health* 33: 164-173.
- Brown, B.B. (1968). *Delphi Process: A Methodology Used for the Elicitation of Opinions of Experts*. Santa Monica, C.A.: The RAND Corporation.
- Byrnes, J.P. (1996). *Cognitive development and learning in instructional contexts*. Boston, MA: Allyn & Bacon.
- Campbell, M.A., & Gardner, S. (2005). A pilot study to assess the effects of life coaching with Year 12 students. In M. Cavanagh, A. Grant, and T. Kemp (Eds.), *Evidence-based coaching*. (pp. 159-169). Brisbane: Australian Academic Press.
- Campbell, N.C., Murray, E., Darbyshire, J., Emery, J., Farmer, A., Griffiths, F., ... Kinmonth, A.L. (2007). Designing and evaluating complex interventions to improve health care. *British Medical Journal* 334(7591): 455-9.
- Carroll, K.M., Nich, C., Sifry, R.L., Nuro, K.F., Frankforter, T.L., Ball, S.A., Fenton, L. & Rounsaville, B.J. (2000). A general system for evaluating therapist adherence and competence in psychotherapy research in addictions. *Drug and Alcohol Dependence* 57: 225-238.
- Cartwright, N., & Hardie, J. (2012). *Evidence-based policy: A practical guide to doing it better*. Oxford University Press.

- Castro, F.P., Barrera, M. & Martinez, C.R. (2004). The cultural adaptation of prevention interventions: resolving tensions between fidelity and fit. *Prevention Science* 5(1): 41-45.
- CBI. (2010). *Ready to grow: business priorities for education and skills*. London: CBI.
- Centre for Contemporary and Cultural Studies (CCCS, 1981). *Unpopular Education: Schooling and Social Democracy in England since 1944*. London: Hutchinson.
- Chalmers, I. & Altman, D.G. (1995). *Systematic Reviews*. London: BMJ.
- Chen, H. (1990). *Theory Driven Evaluations*. Thousand Oaks, CA: Sage Publications.
- Chen, H.T. & Rossi, P.H. (1983). Evaluating With Sense: The Theory-Driven Approach. *Evaluation Review* 7(3): 283-302.
- Chen, H., & Rossi, P.H. (1987). The Theory-Driven Approach to Validity. *Evaluation and Program Planning* 10: 95-103.
- Chitty, C. (2009). *Education Policy in Britain*, 2nd ed. New York: Palgrave Macmillan.
- Cho, H., Hallfors, D. D., & Sánchez, V. (2005). Evaluation of a high school peer group intervention for at-risk youth. *Journal of abnormal child psychology* 33(3): 363-74.
- Christie, C.A. & Alkin, M.C. (2003). The User-Oriented Evaluator's Role in Formulating a Program Theory: Using a Theory-Driven Approach. *American Journal of Evaluation* 24(3): 373-85.
- Connell, J.P. & Kubisch, A.C. (1998). Applying a theory of change approach to the evaluation of comprehensive community initiatives: Progress, prospects and problems. In K. Fullbright-Anderson, A.C. Kubisch & J.P. Connell (Eds.).

- Theory, measurement, and analysis* (Vol. 2). Washington, DC: Aspen Institute.
- The Conservative Party (2010). *Invitation to join the government of Britain. The Conservative manifesto 2010*. London: The Conservative Party.
- Cook, T.D. & Campbell, D.T. (1979). *Quasi-Experimentation: Design and Analysis Issues for Field Settings*. Boston, MA: Houghton Mifflin.
- Cook, T. D., Murphy, R. F., & Hunt, H. D. (2000). Comer's School Development Program in Chicago: A theory-based evaluation. *American Educational Research Journal* 37(2): 535-97.
- Cooksy, L.J., Gill, P. & Kelly, A. (2001). The program logic model as an integrative framework for a multimethod evaluation. *Evaluation and Program Planning* 24: 119-28.
- Cornwall, A., & Jewkes, R. (1995). What is participatory research?. *Social science & Medicine* 41(12): 1667-76.
- Courtney, M.G.R. (2013). Determining the Number of Factors to Retain in EFA: Using the SPSS R-Meny v2.0 to Make More Judicious Estimations. *Practical Assessment, Research & Evaluation* 18(8).
- Cox, C.B. & Boyson, R. (eds) (1975). *Black Paper 1975: The Fight for Education*. London: Dent.
- Cox, C.B. & Boyson, R. (eds) (1977). *Black Paper 1977*. London: Maurice Temple Smith.
- Cox, C.B. & Dyson, A.E. (eds) (1969a). *Black Paper Two: The Crisis in Education*. London: Critical Quarterly Society.
- Cox, C.B. & Dyson, A.E. (eds) (1969b). *Fight for Education: A Black Paper*. London: Critical Quarterly Society.

- Cox C.B. & Dyson, A.E. (eds) (1970). *Black Paper Three: Goodbye Mr Short*. London: Critical Quarterly Society.
- Craig, P., Dieppe, P., Macintyre, S., Michie, S., Nazareth, I. & Petticrew, M. (2008). Developing and evaluating complex interventions: the new Medical Research Council guidance. *BMJ* 337: 979-83.
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological bulletin* 52(4): 281.
- Currie, C., Molcho, M., Boyce, W., Holstein, B., Torsheim, T. & Richter, M. (2008). Researching health inequalities in adolescents: The development of the Health Behaviour in School-Aged Children (HBSC) Family Affluence Scale. *Social Science & Medicine* 66: 1429-36.
- Dalkin, S. M., Greenhalgh, J., Jones, D., Cunningham, B., & Lhussier, M. (2015). What's in a mechanism? Development of a key concept in realist evaluation. *Implementation Science* 10(1): 49.
- Dane, U.A. & Schneider, B.H. (1998). Program integrity in primary and early secondary prevention: are implementation effects out of control? *Clinical Psychology Review* 18(1): 23-45.
- Dariotis, J.K., Bumbarger, B.K., Duncan, L.G. & Greenberg, M.T. (2008). How do implementation efforts relate to program adherence? Examining the role of organizational, implementer, and program factors. *Journal of Community Psychology* 36(6): 744-60.
- Darling-Hammond, L., Ancess, J. & Ort, S.W. (2002). Reinventing high school: outcomes of the coalition campus schools project. *American Educational Research Journal* 39(3): 639-73.

- Deković, M., Asscher, J. J., Manders, W. A., Prins, P. J., & van der Laan, P. (2012). Within-intervention change: mediators of intervention effects during multisystemic therapy. *Journal of consulting and clinical psychology* 80(4): 574.
- Department for Education. (1992). *Choice and Diversity: A New Framework for Schools*, Cmnd 2021. London: HMSO.
- Department for Education. (2010). *NEET Statistics – Quarterly Brief May 2010*. London: Department for Education. Available: <http://www.education.gov.uk/rsgateway/DB/STR/d000924/osr11-2010v2.pdf>
- Department for Education. (2013a). *Free Schools*. Accessed 25 November 2013 at <http://www.education.gov.uk/schools/leadership/typesofschools/freeschools>
- Department for Education (2013b). *What are Studio Schools?* Accessed 12 November 2013 at <http://www.education.gov.uk/schools/leadership/typesofschools/technical/a0077819/about>
- Department for Education (DfE). (2014). *GCSE and equivalent attainment by pupil characteristics*. Accessed 17 March 2016 at <https://www.gov.uk/government/statistics/gcse-and-equivalent-attainment-by-pupil-characteristics-2012-to-2013>.
- Department for Education (DfE). (2015). *Schools, Pupils, and their Characteristics: January 2015*. London: DfE. Available: <https://www.gov.uk/government/statistics/schools-pupils-and-their-characteristics-january-2015>.
- DeVellis, R.F. (2003). *Scale Development: Theory and Applications* (2nd Ed.) Thousand Oaks: Sage.

- Díaz, M., & Simmons, R. (1999). When is research participatory? Reflections on a reproductive health project in Brazil. *Journal of Women's Health*, 8(2): 175-84.
- Dishion, T. J., Li, F., Spracklen, K., Brown, G., & Haas, E. (1998). Measurement of parenting practices in research on adolescent problem behavior: A multimethod and multitrait analysis. *Drug abuse prevention through family interventions*: 260-293.
- Dishion, T. J., McCord, J., & Poulin, F. (1999). When interventions harm: Peer groups and problem behavior. *American psychologist*, 54(9): 755.
- Dobson, D., & Cook, T.J. (1980). Avoiding type III error in program evaluation: Results from a field experiment. *Evaluation and Program Planning* 3: 269-76.
- Domitrovich, C.E. & Greenberg, M.T. (2000). The Study of Implementation: Current Findings from Effective Programs that Prevent Mental Disorders in School-Aged Children. *Journal of Educational and Psychological Consultation* 11(2): 193-221.
- Doward, J. (2011). Government gives go-ahead to first eight “free schools”. *The Guardian*, January 29.
- Dufrene, B.A., Noell, G.H., Gilbertson, D.N. & Duhon, G.J. (2005). Monitoring Implementation of Reciprocal Peer Tutoring: Identifying and Intervening with Students Who Do Not Maintain Accurate Implementation. *School Psychology Review* 34(1): 74-86.
- Durlak, J.A. (1998). Why Program Implementation is Important. *Journal of Prevention and Intervention in the Community* 17(2): 5-18.

- Durlak, J.A. & DuPre, E.P. (2008). Implementation Matters: A Review of Research on the Influence of Implementation on Program Outcomes and the Factors Affecting Implementation. *Am J Community Psychol* 41: 327-50.
- Dusenbury, L., Brannigan, R., Hansen, W.B., Walsh, J. & Falco, M. (2005). Quality of implementation: developing measures crucial to understanding the diffusion of preventive interventions. *Health Education Research* 20(3): 308-13.
- Education Act 2002 (c. 32). London: HMSO.
- Education (Schools) Act 1992 (c.38). London: HMSO.
- Education and Skills Act 2008 (c. 25). London: HMSO.
- Eldridge, S., Spencer, A., Cryer, C., Parsons, S., Underwood, M. & Feder, G. (2005). Why modelling a complex intervention is an important precursor to trial design: lessons from studying an intervention to reduce falls-related injuries in older people. *Journal of Health Services & Policy* 10(3): 133-42.
- Elliott, D.S. & Mihalic, S. (2004). Issues in Disseminating and Replicating Effective Prevention Programs. *Prevention Science* 5(1): 47-53.
- Ertmer, P. A., & Simons, K. D. (2005). Scaffolding teachers' efforts to implement problem-based learning. *International Journal of Learning*, 12(4): 319-328.
- ETS. (2007). *America's Perfect Storm. Three Forces Changing Our Nation's Future*. Princeton, NJ: ETS.
- European Coaching Institute (ECI). (2004). *Benefits of Coaching in Business*. Weymouth: ECI Accreditation Limited.
- Evans, R., Scourfield, J., & Murphy, S. (2014). Pragmatic, formative process evaluations of complex interventions and why we need more of them. *Journal of epidemiology and community health* 1-2.

- Fagan, A.A., Hanson, K., Hawkins, J.D. & Arthur, M.W. (2008). Bridging Science to Practice: Achieving Prevention Program Implementation Fidelity in the Community Youth Development Study. *American Journal of Community Psychology* 41: 235-49.
- Feldman, D.C. & Lankau, M.J. (2005). Executive Coaching: A Review and Agenda for Future Research. *Journal of Management* (31): 829-48.
- Feinstein, L. (2002). *Quantitative Estimates of the Social Benefits of Learning, 1: Crime. Wider Benefits of Learning Research Report No. 5*. London: Centre for Research on the Wider Benefits of Learning.
- Fine, M. (1994). *Chartering urban school reform*. In M. Fine (Ed.), *Chartering urban school reform* (pp. 5–30). New York, NY: Teachers College Press.
- Fitch, K., Bernstein, S.J., Aguilar, M.D., Burnand, B., LaCalle, J.R., Lazaro, P., van het Loo, M., McDonnell, J., Vader, J.P. & Kahan, J.P. (2001). *The RAND/UCLA Appropriateness Method User's Manual*. Santa Monica, CA: RAND Corporation.
- Forgatch, M. S., Patterson, G. R., & DeGarmo, D. S. (2006). Evaluating fidelity: Predictive validity for a measure of competent adherence to the Oregon model of parent management training. *Behavior therapy* 36(1): 3-13.
- Franklin, B. & Crone, L. (1992). *School accountability: Predictors and indicators of Louisiana school effectiveness*. Paper presented at the meeting of the Mid-South Educational Research Association, Knoxville, TN.
- Fraser, M.W., Richman, J.M., Galinsky, M.J. & Day, S.H. (2009). *Intervention Research: Developing Social Programs*. Oxford: Oxford University Press.

- Fullarton, S. (1999). *Work experience and work placements in secondary school education*. Camberwell, Victoria: Australian Council for Education Research.
- Gardner, B., Whittington, C., McAteer, J., Eccles, M.P. & Michie, S. (2010). Using theory to synthesise evidence from behaviour change interventions: The example of audit and feedback. *Social Science & Medicine* 70: 1618-25.
- Gardner, F., Hutchings, J., Bywater, T., & Whitaker, C. (2010). Who benefits and how does it work? Moderators and mediators of outcome in an effectiveness trial of a parenting intervention. *Journal of Clinical Child & Adolescent Psychology*, 39(4), 568-580.
- Gibbs, L. & Gambrill, E. (2002). Evidence-Based Practice: Counterarguments to Objections. *Research on Social Work Practice* 12: 452-75.
- Gijpels, D., Dochy, F., Van den Bossche, P. & Segers, M. (2005). Effects of Problem-Based Learning: A Meta-analysis from the Angle of Assessment. *Review of Educational Research* 75(1): 27-61.
- Gillard, D. (2011). *Education in England: a brief history*. Accessed 12 November 2013 at <http://www.educationengland.org.uk/history/index.html>.
- Glanz, K., & Bishop, D.B. (2010). The role of behavioral science theory in development and implementation of public health interventions, *Annual Review of Public Health* 31: 399-418.
- Glasgow, R.E., Klesges, L.M., Dzewaltowski, D., Bull, S.S. & Estabrooks, P. (2004). The future of health behavior change research: What is needed to improve translation of research into health promotion practice? *Annals of Behavioral Medicine* 27: 3-12.
- Glennerster, H. (2000). *British Social Policy since 1945*. Oxford: Blackwell Publishing.

- Glouberman, S., & Zimmerman, B. (2002). Complicated and complex systems: what would successful reform of Medicare look like? *Romanow Papers* 2: 21-53.
- Gottfredson, D.C. & Gottfredson, G.D. (2002). Quality of school-based prevention programs: Results from a national survey. *Journal of Research in Crime and Delinquency* 39: 3-35.
- Grant, S.P. (2014). Development of a CONSORT extension for social and psychological interventions. Unpublished thesis. Available: <http://solo.bodleian.ox.ac.uk>.
- Grant, S., Mayo-Wilson, E., Hopewell, S., Macdonald, G., Moher, D., & Montgomery, P. (2013). Developing a reporting guideline for social and psychological intervention trials. *Journal of Experimental Criminology* 9(3): 355-67.
- Green, L. W., & Mercer, S. L. (2001). Can public health researchers and agencies reconcile the push from funding bodies and the pull from communities?. *American journal of public health* 91(12): 1926-9.
- Gresham, F.M., Gansle, K.A. & Noell, G.H. (1993). Treatment Integrity in Applied Behavior Analysis with Children. *Journal of Applied Behavior Analysis* 26(2): 257-63.
- Griffiths, K. (2005). Personal Coaching: A Model for Effective Learning. *Journal of Learning Design* 1(2): 55-65.
- Hahn, E.J., Noland, M.P., Rayens, M.K. & Christie, D.W. (2002). Efficacy of Training and Fidelity of Implementation of the Life Skills Training Program. *Journal of School Health* 72(7): 282-7.
- Hair, J., Anderson, R.E., Tatham, R.L., & Black, W.C. (1995). *Multivariate data analysis: 4th ed.* New York: Prentice-Hall Inc.

- Hamilton, M.A. & Hamilton, S.F. (1997). *Learning Well at Work: Choices for Quality*. Ithaca, NY: Cornell University.
- Hansen, W. B., Graham, J. W., Wolkenstein, B. H., & Rohrbach, L. A. (1991). Program integrity as a moderator of prevention program effectiveness: Results for fifth grade students in the adolescent alcohol prevention trial. *Journal of Studies on Alcohol* 52: 568–579.
- Hardeman, W., Michie, S., Fanshawe, T., Prevost, A. T., Mcloughlin, K., & Kinmonth, A. L. (2008). Fidelity of delivery of a physical activity intervention: predictors and consequences. *Psychology and Health* 23(1): 11-24.
- Hargreaves, J., Hatcher, A., Strange, V., Phetla, G., Busza, J., Kim, J., ... & Bonell, C. (2009). Process evaluation of the Intervention with Microfinance for AIDS and Gender Equity (IMAGE) in rural South Africa. *Health Education Research* 25(1): 27-40.
- Hasson, F., Keeney, S. & McKenna, H. (2000). Research guidelines for the Delphi survey technique. *Journal of Advanced Nursing* 32(4): 1008-15.
- Hatcher, R. (2011). The Conservative-Liberal Democrat Coalition government’s “free schools” in England. *Educational Review* 63(4): 485-503.
- Hawe, P., Shiell, A. & Riley, T. (2004a). Complex interventions: how out-of-control can a randomised controlled trial be? *BMJ* 328: 1561-3.
- Hawe, P., Shiell, A., Riley, T., & Gold, L. (2004b). Methods for exploring implementation variation and local context within a cluster randomised community intervention trial. *Journal of Epidemiology and Community Health* 58(9): 788-93.

- Hawe, P., Shiell, A. & Riley, T. (2009). Theorizing interventions as Events in Systems. *American Journal of Community Psychology* 43: 267-76.
- Hayduk, L., Cummings, G., Boadu, K., Pazderka-Robinson, H & Boulianne, S. (2007). Testing? Testing! one, two three – Testing the theory in structural equation models! *Personality and Individual Differences* 42: 841-50.
- Heckman, J.J. & Rubinstein Y. (2001). The Importance of Noncognitive Skills: Lessons Learned from the GED Testing Program. *The American Economic Review* 91(2): 145-9.
- Heckman, J.J. (2006). Skill Formation and the Economics of Investing in Disadvantaged Children. *Science* 312: 1900-1902.
- Henngeler, S.W., Schoenwald, S.K., Borduin, C.M. & Swenson, C.C. (2006). Methodological critique and meta-analysis as Trojan horse. *Children and Youth Services Review* 28: 447-457.
- Hershey, A.M., Silverberg, M.K., Haimson, J., Hudis, P. & Jackson, R. (1999). *Expanding options for students: Report to Congress on the national evaluation of school-to-work implementation*. Princeton, NJ: Mathematica Policy Research, Inc.
- Hill, L.G., Maucione, K. & Hood, B.K. (2007). A Focused Approach to Assessing Program Fidelity. *Prevention Science* 8: 25-34.
- Hogendoorn, S. M., Prins, P. J., Boer, F., Vervoort, L., Wolters, L. H., Moorlag, H., ... & de Haan, E. (2014). Mediators of cognitive behavioral therapy for anxiety-disordered children and adolescents: Cognition, perceived control, and coping. *Journal of Clinical Child & Adolescent Psychology* 43(3): 486-500.

- Hollenbeck, K. (1995). *In their own words: Student perspectives on school-to-work opportunities*. National Institute for Work and Learning & Academy for Educational Development.
- Hopewell, S., Clarke, M., Moher, D., Wager, E., Middleton, P., Altman, D. G., ... & CONSORT Group. (2008). CONSORT for reporting randomized controlled trials in journal and conference abstracts: explanation and elaboration. *PLoS Med*, 5(1): e20.
- Howieson, C. (1990). Beyond the Gate: Work Experience and Part-Time Work among Secondary-School Pupils in Scotland. *British Journal of Education and Work* 3(3): 49-61.
- Hughes, K.L., Moore, D.T. & Bailey, T.R. (1999). *Work-Based Learning and Academic Skills*. IEE Working Paper No. 15.
- Huhman, M., Heitzler, C., & Wong, F. (2004). The VERB™ campaign logic model: a tool for planning and evaluation. *Preventing chronic disease* 1(3): 1-6.
- Julian, D. A. (1997). The utilization of the logic model as a system level planning and evaluation device. *Evaluation and program planning* 20(3), 251-7.
- Jagosh, J., Pluye, P., Macaulay, A. C., Salsberg, J., Henderson, J., Sirett, E., ... & Cargo, M. (2011). Assessing the outcomes of participatory research: protocol for identifying, selecting, appraising and synthesizing the literature for realist review. *Implementation Science* 6(1): 24.
- Kam, C. M., Greenberg, M. T., & Walls, C. T. (2003). Examining the role of implementation quality in school-based prevention using the PATHS curriculum. *Prevention Science* 4(1): 55-63.
- Katz, J.H. & Miller, F.A. (1996). Coaching Leaders Through Culture Change. *Consulting Psychology Journal: Practice and Research* 48(2): 104-114.

- Kellogg Foundation, W.K. (2004). *Logic model development guide*. Battle Creek, MI: WK Kellogg Foundation.
- Kemple, J.J. & Snipes, J.C. (2000). *Career Academies: Impact on Students' Engagement and Performance in High School*. New York, NY: Manpower Demonstration Research Corporation.
- Kennedy, H.P. (2004). Enhancing Delphi research: methods and results. *Journal of Advanced Nursing* 45(5): 504-11.
- Kilburg, R.R. (1996). Toward a Conceptual Understanding and Definition of Executive Coaching. *Consulting Psychology Journal: Practice and Research* 48(2): 134-44.
- Kinmonth, A. L., Wareham, N. J., Hardeman, W., Sutton, S., Prevost, A. T., Fanshawe, T., ... & Griffin, S. J. (2008). Efficacy of a theory-based behavioural intervention to increase physical activity in an at-risk group in primary care (ProActive UK): a randomised trial. *The Lancet* 371(9606): 41-48.
- Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational psychologist* 41(2): 75-86.
- Kline, R.B. (2011). *Principles and Practice of Structural Equation Modeling* (3rd Ed.). New York: The Guilford Press.
- Kolodner, J. L., Camp, P. J., Crismond, D., Fasse, B., Gray, J., Holbrook, J., ... & Ryan, M. (2003). Problem-based learning meets case-based reasoning in the middle-school science classroom: Putting learning by design (tm) into practice. *The journal of the learning sciences* 12(4): 495-547.

- Krieger, J., Allen, C., Cheadle, A., Ciske, S., Schier, J. K., Senturia, K., & Sullivan, M. (2002). Using community-based participatory research to address social determinants of health: lessons learned from Seattle Partners for Healthy Communities. *Health Education & Behavior*, 29(3): 361-82.
- Kumpfer, K.L., Alvarado, R., Smith, P. & Bellamy, N. (2002). Cultural sensitivity and adaption in family-based prevention interventions. *Prevention Science* 3: 241-246.
- Lacouture, A., Breton, E., Guichard, A. & Ridde, V. (2015). The concept of mechanism from a realist approach: a scoping review to facilitate its operationalisation in public health program evaluation. *Implementation Science*. DOI: 10.1186/s13012-015-0345-7,
- Lane, K.L., Bocian, K.M., MacMillan, D.L. & Gresham, F.M. (2004). Treatment Integrity: An Essential -But Often Forgotten- Component of School-Based Interventions. *Preventing School Failure* 48(3): 36-43.
- Lapan, R. T., Gysbers, N. C., & Petroski, G. F. (2001). Helping seventh graders be safe and successful: A statewide study of the impact of comprehensive guidance and counseling programs. *Journal of Counseling and Development* 79(3): 320.
- Lawton, D. (ed) (1989). *The Education Reform Act: Choice and Control*. London: Hodder & Stoughton.
- Learning and Skills Act 2000 (c. 21). London: HMSO.
- Lee, S.J., Altschul, I. & Mowbray, C.T. (2008). Using Planned Adaptation to Implement Evidence-Based Programs with New Populations. *American Journal of Community Psychology* 41: 290-303.

- Leonard, M. (1988). *The 1988 Education Act: A Tactical Guide for Schools*. Oxford: Basil Blackwell Ltd.
- Lillehoj, C.J., Griffin, K.W., & Spoth, R. (2004). Program provider and observer ratings of school- based preventive intervention implementation: Agreement and relation to youth outcomes. *Health Education and Behavior* 31: 242–257.
- Lipsey, MW. (1999). Can Intervention Rehabilitate Serious Delinquents? *Annals of the American Academy of Political and Social Science* 564: 142-66.
- Littell, J.H., Popa, M. & Forsythe, B. (2005). Multisystemic Therapy for social, emotional, and behavior problems in youth age 10-17. *Cochrane library, issue 3, 2005*. Chichester: John Wiley & Sons, Ltd.
- Livet, M., & Wandersman, A. (2005). Organizational functioning: Facilitating effective interventions and increasing the odds of programming success. *Empowerment evaluation principles in practice* 8: 123-54.
- Lleras, C. (2008). Do skills and behaviors in high school matter? The contribution of noncognitive factors in explaining differences in educational attainment and earnings. *Social Science Research* 37: 888-902.
- Lochner, L. & Moretti, E. (2001). *The Effect of Education on Crime: Evidence from Prison Inmates, Arrests and Self-Reports*. National Bureau of Economic Research Working Paper 8605.
- Lorencatto, F., West, R., Stavri, Z., & Michie, S. (2012). How well is intervention content described in published reports of smoking cessation interventions? *Nicotine & Tobacco Research*, 15(7): 1273-82.
- Machin, M.A. & Fogarty, G.J. (2003). Perceptions of training-related factors and personal variables as predictors of transfer implementation intentions. *Journal of Business and Psychology* 18(1): 51-71.

- Manlove, J. (1998). The Influence of High School Dropout and School Disengagement on the Risk of School-Age Pregnancy. *Journal of Research on Adolescence* 8(2): 187-220.
- Marshall, P. (ed.). (2013). *The Tail: How England's schools fail one child in five – and what can be done*. London: Profile Books.
- McCulloch, G. (1994). *Educational Reconstruction: The 1944 Education Act and the Twenty-first Century*. Ilford: The Woburn Press.
- McHugh, R.K., Murray, H.W. & Barlow, D.H. (2009). Balancing fidelity and adaptation in the dissemination of treatments: The promise of transdiagnostic interventions. *Behaviour Research and Therapy* 47: 946-53.
- McIntosh, S. and Houghton, N. (2005) *Disengagement from Secondary Education: A story retold*. London: Learning and Skills Network.
- McKenna, H. P. (1994). The Delphi technique: a worthwhile research approach for nursing? *Journal of advanced nursing* 19(6): 1221-5.
- McKinney, E. (2015). Two Midland ‘revolutionary’ studio schools to close. *Birmingham Post*, 10 Dec 2015. Available: <http://www.birminghampost.co.uk/news/regional-affairs/two-midland-revolutionary-studio-schools-10580592>.
- McLaughlin, J. A., & Jordan, G. B. (1999). Logic models: a tool for telling your programs performance story. *Evaluation and program planning* 22(1): 65-72.
- Michie, S., & Abraham, C. (2004). Identifying techniques that promote health behaviour change: Evidence based or evidence inspired. *Psychol Health* 19: 29-49.
- Michie, S., Johnston, M., Abraham, C., Lawton, R., Parker, D., & Walker, A. on behalf of the “Psychological Theory” Group. (2005). Making psychological

- theory useful for implementing evidence based practice: a consensus approach. *Quality and Safety in Health Care* 14: 26-33.
- Michie, S., Johnston, M., Francis, J., Hardeman, W., & Eccles, M. (2008a). From theory to intervention: mapping theoretically derived behavioural determinants to behaviour change techniques. *Applied psychology* 57(4): 660-680.]
- Michie, S., Hardeman, W., Fanshawe, T., Prevost, A. T., Taylor, L., & Kinmonth, A. L. (2008). Investigating theoretical explanations for behaviour change: The case study of ProActive. *Psychology and Health* 23(1): 25-39.
- Michie, S. & Prestwich, A. (2010). Are Interventions Theory-Based? Development of a Theory Coding Scheme. *Health Psychology* 29(1): 1-8.
- Mihalic, S. (2004). The importance of implementation fidelity. *Emotional and Behavioral Disorders in Youth* 4(4): 83–105.
- Mizen, P. (2004). *The changing state of youth*. New York, NY: Palgrave Macmillan.
- Monroe, M. C., Fleming, M. L., Bowman, R. A., Zimmer, J. F., Marcinkowski, T., Washburn, J., & Mitchell, N. J. (2005). Evaluators as educators: Articulating program theory and building evaluation capacity. *New Directions for Evaluation* 108: 57-71.
- Montgomery, P., Underhill, K., Gardner, F., Operario, D. & Mayo-Wilson, E. (Expected: 2013). The Oxford Implementation Index: A new tool for incorporating implementation data into systematic reviews and meta-analyses. *Journal of Clinical Epidemiology* 66(8): 874-82.
- Moore, G., Audrey, S., Barker, M., Bond, L., Bonell, C., Hardeman, W., Moore, L., O’Cathain, A., Tinati, T., Wight D. & Baird, J. (2014). *Process evaluation of complex interventions: Medical Research Council guidance*. London: MRC Population Health Research Network.

- Moore, J.E. Bumbarger, B.K. & Cooper, B.R. (2013). Examining Adaptations of Evidence-Based Programs in Natural Contexts. *Journal of Primary Prevention*.
- Morrison Gutman, L., Brown, J., Akerman, R. & Obolenskaya, P. (2010). *Change in wellbeing from childhood to adolescence: risk and resilience. Research Report 34*. London: CRWBL.
- Moursund, D. (2003). *Project-based learning using information technology*. Eugene, OR: International Society for Technology in Education.
- Mowbray, C.T., Holter, M.C., Teague, G.B. & Bybee, D. (2003). Fidelity Criteria: Development, Measurement, and Validation. *American Journal of Evaluation* 24: 315-40.
- Mullen, E.J., Bledsoe, S.E. & Bellamy, J.L. (2008). Implementing Evidence-Based Social Work Practice. *Research on Social Work Practice* 18: 325-38.
- Murphy, MK., Sanderson, C.F.B., Black, N.A., Ashkam, J., Lamping, D.L., Marteau, T. & McKee, C.M. (1998). Consensus development methods, and their use in clinical guideline development. *Health Technology Assessment* 2(3).
- National Institutes of Health. (2008). *Community participation in research. NIH guide to grants*. Available: <http://grants.nih.gov/grants/guide/pa-files/PAR-05-026.html>
- National Youth Employment Council (NYEC, 1974). *Unqualified, Untrained and Unemployed*. London: HMSO.
- Newman, M., Garrett, Z., Elbourne, D., Bradley, S., Noden, P., Taylor, J. & West, A. (2006). Does secondary school size make a difference? A systematic review. *Educational Research Review* 1: 41-60.

- Noar, S.M., Chabot, M., and Zimmerman, R.S. (2008). Applying health behavior theory to multiple behavior change: considerations and approaches. *Preventive Medicine* 46(3): 275-80.
- Noar, S. M., & Zimmerman, R. S. (2005). Health Behavior Theory and cumulative knowledge regarding health behaviors: are we moving in the right direction? *Health education research* 20(3): 275-90.
- Oakley, A., Strange, V., Bonell, C., Allen, E., Stephenson, J. & the RIPPLE Study Team. (2006). Process evaluation in randomized controlled trials of complex interventions. *BMJ* 332: 413-6.
- OECD, (2013a). *PISA: Snapshot of performance in mathematics, reading and science*. OECD Publishing.
- OECD. (2013b). *OECD Factbook 2013: Economic Environmental and Social Statistics*. OECD Publishing.
- Intervention, n. *OED Online*. Oxford: Oxford University Press, December 2015.
- Office for National Statistics. (2014a). *School workforce in England: November 2014*. Available: <https://www.gov.uk/government/statistics/school-workforce-in-england-november-2014>
- Office for National Statistics. (2014b). *LI01: Local labour market indicators by Unitary and Local Authority*. Available: <http://www.ons.gov.uk/ons/taxonomy/search/index.html?nscl=Unemployment+by+Region&nscl-orig=Unemployment+by+Region&content-type=Dataset&contenttype=Reference+table&sortDirection=DESCENDING&sortBy=pubdate>.
- OFSTED. (2015a). *Inspection report: Stephenson Studio School*. Manchester: OFSTED.

- OFSTED. (2015b). *Inspection report: Waverley Studio College*. Manchester: OFSTED.
- OFSTED. (2015c). *Inspection report: Darwen Aldridge Enterprise Studio*. Manchester: OFSTED.
- OFSTED. (2015d). *Inspection report: Rye Studio School*. Manchester: OFSTED.
- Ogilvie, D., Cummins, S., Petticrew, M., White, M., Jones, A. & Wheeler, K. (2011). Assessing the Evaluability of Complex Public Health Interventions: Five Questions for Researchers, Funders, and Policymakers. *The Milbank Quarterly* 89(2): 206-25.
- Okoli, C. & Pawlowski, S.D. (2004). The Delphi method as a research tool: an example, design considerations and applications. *Information & Management* 42(1): 15-29.
- Oosthuizen, C. & Louw, J. (2013). Developing program theory for purveyor programs. *Implementation Science* 8(23).
- Oxley, D. (1995). *The making of a school community: Organizational structure, processes, and goals*. In Deffenbaug, A. (Ed.), *The Oregon conference monograph* 7 (pp. 43–56). Eugene, OR: Eugene College of Education.
- Ozer, E.J., Wanis, M.G. & Bazell, N. (2010). Diffusion of School-Based Prevention Programs in Two Urban Districts: Adaptations, Rationales, and Suggestions for Change. *Prevention Science* 11: 42-55.
- Pagel, M. (2012). *Wired for culture: origins of the human social mind*. WW Norton & Company.
- Patton, M.Q. (2008). *Utilization-focused evaluation*. Thousand Oaks: Sage.
- Pawson, R. (2013). *The Science of Evaluation: A Realist Manifesto*. Thousand Oaks: Sage.

- Pawson, R. & Tilley, N. (1997). *Realistic Evaluation*. London: Sage Publications Ltd.
- Pawson, R. & Tilley, N. (2004). *Realist Evaluation*. Available:
http://www.communitymatters.com.au/RE_chapter.pdf.
- Pereplechikova, F. & Kazdin, A.E. (2005). Treatment Integrity and Therapeutic Change: Issues and Research Recommendations. *Clinical Psychology: Science and Practice* 12(4): 365-383.
- Pereplechikova, F., Treat, T. A., & Kazdin, A. E. (2007). Treatment integrity in psychotherapy research: analysis of the studies and examination of the associated factors. *Journal of consulting and clinical psychology*, 75(6): 829.
- Pollard, C., & Pollard, R. (2004). Research priorities in educational technology: A Delphi study. *Journal of Research on Technology in Education* 37(2): 145-60.
- Powell, C. (2002). The Delphi technique: myths and realities. *Journal of Advanced Nursing* 41(4): 376-82.
- Pornprasertmanit, S., Lee, J., & Preacher, K. J. (2014). Ignoring clustering in confirmatory factor analysis: some consequences for model fit and standardized parameter estimates. *Multivariate behavioral research*, 49(6): 518-543.
- Preacher, K. J., Rucker, D. D., & Hayes, A. F. (2007). Addressing moderated mediation hypotheses: Theory, methods, and prescriptions. *Multivariate behavioral research* 42(1): 185-227.
- Pring, R. & Walford, G. (1996). *Comprehensive schools: the history*. Times Higher Education. Available:
<https://www.timeshighereducation.com/news/comprehensive-schools-the-history/92186.article>

- Ravitz, J. (2008). *Project Based Learning as a Catalyst in Reforming High Schools*. New York, NY: AERA.
- Raywid, M.A. (1998). Small Schools: A Reform that Works. *Educational Leadership* 55(4), 34–39.
- Renger, R. (2010). Constructing and verifying program theory using source documentation. *The Canadian Journal of Program Evaluation* 25(1): 51.
- Resnicow, K., Davis, M., Smith, M., Lazarus-Yaroch, A., Baranowski, T., Baranowski, J., et al. (1998). How best to measure implementation of school health curricula: A comparison of three measures. *Health Education Research* 13: 239–250.
- Resnicow, K., Soler, R., Braithwaite, R.L., Ahluwalia, J.S. & Butler, J. (2000). Cultural Sensitivity in Substance Abuse Prevention. *Journal of Community Psychology* 28(3): 271-90.
- Richardson, W. & Wiborg for the Baker Dearing Educational Trust. (2010). *English Technical and Vocational Education in Historical and Comparative Perspective*. London: Baker Dearing Educational Trust.
- Ringwalt, C.L., Pankratz, M.L., Hansen, W.B., Dusenbury, L., Jaskon-Newsom, J., Giles, S.M. & Brodish, P.H. (2009). The Potential of Coaching as a Strategy to Improve the Effectiveness of School-Based Substance Use Prevention Curricula. *Health Education & Behavior* 36: 696-710.
- Ringwalt, C.L., Vincus, A., Ennett, S., Johnson, R. & Rohrbach, L.A. (2004). Reasons for Teachers' Adaptation of Substance Use Prevention Curricula in Schools With Non-White Student Populations. *Prevention Science* 5(1): 61-67.
- Rogers, E.M. (2003). *Diffusion of Innovations (5th ed)*. New York, NY: Free Press.

- Rogers, P.J. (2007). Theory-Based Evaluation: Reflections Ten Years On. *New Direction for Evaluation* 114: 63-81.
- Ross, A. (2009). *Disengagement from Education among 14-16 year olds*. London: Department for Children, Schools and Families.
- Rothstein, R. (2004). *Class and Schools: Using Social, Economic, and Educational Reform to Close the Black-White Achievement Gap*. Washington, DC: Economic Policy Institute.
- Ryan, P. (2001). The School-to-Work Transition: A Cross-National Perspective. *Journal of Economic Literature* 39(1): 34-92.
- Sackett, D.L., Rosenberg, W.M., Gray, J.A.M., Haynes, R.B. & Richardson, W.S. (1996). Evidence-based medicine: what it is and what it isn't. *BMJ* 312: 71-72.
- Sandbrook, D. (2006). *White Heat: A History of Britain in the Swinging Sixties*. London: Little, Brown Group.
- Sandbrook, D. (2012). *Seasons in the Sun: The Battle for Britain, 1974-1979*. London: Penguin Books Ltd.
- Sanders, M.R. (1999). Triple P-Positive Parenting Program: Towards an Empirically Validated Multilevel Parenting and Family Support Strategy for the Prevention of Behavior and Emotional Problems in Children. *Clinical Child and Family Psychology Review* 2(2): 71-90.
- Saunders, R. P., Ward, D., Felton, G. M., Dowda, M., & Pate, R. R. (2006). Examining the link between program implementation and behavior outcomes in the lifestyle education for activity program (LEAP). *Evaluation and program planning* 29(4): 352-64.

- Savaya, R., & Waysman, M. (2005). The logic model: A tool for incorporating theory in development and evaluation of programs. *Administration in Social Work* 29(2): 85-103.
- Schank, R.C., Fano, A., Bell B. & Jona, M. (1994). The Design of Goal-Based Scenarios. *Journal of the Learning Sciences*, 3(4): 305-45.
- Schinke, S., Brounstein, P. & Gardner, S. (2002). *Science-based prevention programs and principles*, 2002, DHHS Pub No. (SMA) 03-3764, Center for Substance Abuse and Mental Health Services Administration, Rockville, MD.
- Shekelle, P. (2004). The Appropriateness Method. *Medical Decision Making* 228-31.
- Sinha, I. P., Smyth, R. L., & Williamson, P. R. (2011). Using the Delphi technique to determine which outcomes to measure in clinical trials: recommendations for the future based on a systematic review of existing studies. *PLoS Med* 8(1): e1000393.
- Solomon, G. (2003). Project-Based learning: A Primer. *Technology and Learning* 23(6): 20–30.
- Spiro, R.J., Coulson, R.L., Feltovich, P.J. & Anderson, D.K. (1988). *Cognitive Flexibility Theory: Advanced Knowledge Acquisition in Ill-Structured Domains*. In: V. Patel (Eds.), Tenth Annual Conference of the Cognitive Science Society Proceedings (pp. 640-653). Hillsdale, NJ: Lawrence Erlbaum.
- Stasz, C. & Brewer, D.J. (1999). *Academic Skills at Work: Two Perspectives*. Berkeley, CA: National Center for Research in Vocational Education.
- Steedman, H. & Stoney, S. (2004). *Disengagement 14-16: Context and Evidence*. CEP Discussion Paper No. 654. London: London School of Economics/CEP.
Available: <http://cep.lse.ac.uk/pubs/download/dp0654.pdf>

- Steifel, L., Iatarola, P., Fruchter, N., & Berne, R. (1998). *The effects of size of student body on school costs and performance in New York City high schools*. New York, NY: Institute for Education and Society Policy. Available: www.nyu.edu/iesp/publications/effects/effects.pdf
- Stewart, R.E. & Chambless, D.L. (2007). Does Psychotherapy Research Inform Treatment Decisions in Private Practice? *Journal of Clinical Psychology* 63: 267-81.
- Studio Schools Trust (2012a). *The Studio Schools Model: an Introduction*. Manchester: Studio Schools Trust. Available: <http://www.onderwijsvanmorgen.nl/assets/Studio-Schools-Trust-brochure.pdf>
- Studio Schools Trust (2012b) *The need for Studio Schools: Overview of the Evidence*. Manchester: Studio Schools Trust. Available: <http://new.studioschoolstrust.org/sites/default/files/Studio%20Schools%20Research%20Report.pdf>
- Studio Schools Trust (2013a). *The Trust*. Accessed 12 November 2013 at <http://www.studioschoolstrust.org/the-trust>
- Studio Schools Trust (2013b). *Essential Elements of a Studio School*. Accessed 12 November 2013 at <http://www.studioschoolstrust.org/studio-schools/essential-elements-studio-school>
- Studio Schools Trust (2013c). *CREATE Framework*. Accessed 12 November 2013 at
- Sullivan, H., Barnes, M., & Matka, E. (2002). Building Collaborative Capacity through Theories of Change. Early Lessons from the Evaluation of Health Action Zones in England. *Evaluation* 8(2), 205-26.
- SurveyMonkey Inc. (2013). Survey Monkey [Software]. Palo Alto, California, USA, www.surveymonkey.com

- Thomas, J.W. (2000). *A Review of Research on Project-Based Learning*. San Rafael, CA: The Autodesk Foundation.
- Thomas, W. & Smith, A. (2009). *Coaching Solutions. Practical Ways to Improve Performance in Education* (2nd ed). London: Continuum International Publishing Group.
- Thompson, B., Coronado, G., Snipes, S. A., & Puschel, K. (2003). Methodologic advances and ongoing challenges in designing community-based health promotion programs. *Annual Review of Public Health* 24(1): 315-40.
- To, T., Guttman, A., Lougheed, M.D., Gershon, A.S., Dell, S.D., Stanbrook, M.B., Wang, C., McLimont, S., Vasilevska-Ristovska, J., Crighton, E.J. & Fisman, D.N. (2010). Evidence-based performance indicators of primary care for asthma: a modified RAND Appropriateness Method. *International Journal for Quality in Health Care* 22(6): 476-85.
- Trevisan, M.S. (2007). Evaluability Assessment From 1986 to 2006. *American Journal of Evaluation* 28: 290-303.
- Trifiletti, L. B., Gielen, A. C., Sleet, D. A., & Hopkins, K. (2005). Behavioral and social sciences theories and models: are they used in unintentional injury prevention research?. *Health Education Research* 20(3): 298-307.
- Van Urk, F.C., Grant, S., & Bonell, C. (2015). Involving Stakeholders in Programme Theory Specification: Discussion of a Systematic, Consensus-Based Approach. *Evidence & Policy*. DOI: <http://dx.doi.org/10.1332/174426415X14474260456850>.
- Veenman, S., Gerrits, J. & Kenter, J. (1999). Coaching pre-service teachers. In *Abstracts 8th European Conference on Learning and Instruction* (p.548).

- Goteborg, Sweden: European Association for Research on Learning and Instruction.
- Vernon, D.T.A. & Blake, R.L. (1993). Does Problem-Based Learning Work? A Meta-analysis of Evaluative Research. *Academic Medicine* 68(7): 550-63.
- Walker, A. & Leary, H. (2009). A Problem Based Learning Meta Analysis: Differences Across Problem Types, Implementation Types, Disciplines, and Assessment Levels. *Interdisciplinary Journal of Problem-based Learning* 3(1): 6-28.
- Wandersman, A. (2003). Community science: Bridging the gap between science and practice with community-centered models. *American Journal of Community Psychology* 31: 227-42.
- Wandersman, A., Duffy, J., Flashpohler, P., Noonan, R., Lubell, K., Stillman, L., Blachman, M., Dunville, R. & Saul, J. (2008). Bridging the Gap Between Prevention Research and Practice: The Interactive Systems Framework for Dissemination and Implementation. *American Journal of Community Psychology* 41: 171-81.
- Wasley, P.A. & Lear, R.J. (2001). Small Schools, Real Gains. *Educational Leadership Magazine* 58(6).
- Webster-Stratton, C. (2004). *Quality training, supervision, ongoing monitoring, and agency support: key ingredients to implementing the incredible years programs with fidelity*. Treatment description, University of Washington.
- Weiss, C.H. (1995). Nothing as Practical as Good Theory. In J. Connell, A. Kubisch, L.B. Schorr & C.H. Weiss (eds.). *New Approaches to Evaluating Community Initiatives*. New York, NY: Aspen Institute.

- Weiss, C.H. (1997). Theory-Based Evaluation: Past, Present, and Future. In D. Rog and D. Founier (eds.), *Progress and Future Directions in Evaluation: Perspectives on Theory, Practice, and Methods*. New Directions for Program Evaluation, no. 76. San Francisco: Jossey-Bass.
- Weiss, C.H. (2004). Which Links in Which Theories Shall We Evaluate? *New Directions for Evaluation* 87: 35-45.
- Wholey, J. S. (1996). Formative and summative evaluation: Related issues in performance measurement. *American Journal of Evaluation* 17(2): 145-149.
- Williams, R.P. & Yeomans, D.J. (1994). The Technical and Vocational Education Initiative and school autonomy in the management of curriculum change. *Research Papers in Education* 9(3): 303-19.
- Williams, B., Brown, T. & Onsmann, A. (2012). Exploratory factor analysis: A five-step guide for novices. *Australasian Journal of Paramedicine* 8(3): 1-13.
- Wilson, D.B., MacKenzie, D.L. & Mitchell, F.N. (2003). Effects of correctional boot camps on offending. *Campbell Systematic Reviews* 2003: 1. DOI: 10.4073/csr.2003.1.
- Wolf, A. (2002). *Does education matter? Myths about education and economic growth*. London: Penguin Books Ltd.
- Working Group on 14-19 Reform. (2004). *14-19 Curriculum and Qualifications Reform: Final Report of the Working Group on 14-19 Reform*. Annesley: DfES Publications.
- Yampolskaya, S., Nesman, T. M., Hernandez, M., & Koch, D. (2004). Using concept mapping to develop a logic model and articulate a program theory: A case example. *American Journal of Evaluation* 25(2): 191-207.

Zeus, P., & Skiffington, S. (2002). *The coaching at work toolkit: A complete guide to techniques and practices*. Sydney: McGraw Hill.

Function-focused implementation fidelity for complex interventions: the case of Studio Schools

Appendices

Felix Constantijn van Urk, MSc



Appendices: Table of contents

Appendix A: Delphi process materials and supplements	2
<i>A1: Questionnaire used for Delphi Round One</i>	2
<i>A2: Original logic model items nominated by participants in Delphi Round One</i>	7
<i>A3: List of non-redundant items, resulting from integration of nominated items by the researchers (rated by participants in Delphi Round Two)</i>	31
<i>A4: Results of Delphi Round Two</i>	38
<i>A5: Decision process on items classified as uncertain after Delphi Round Two</i>	58
 Appendix B: Implementation checklists and questionnaires	 63
<i>B1: Checklist used by researcher for rating PBL implementation</i>	63
<i>B2: Questionnaire used by students for rating PBL implementation</i>	66
<i>B3: Questionnaire used by teachers for rating PBL implementation</i>	71
<i>B4: Checklist used by researcher for rating personal coaching implementation</i>	76
<i>B5: Questionnaire used by students for rating personal coaching implementation</i>	79
<i>B6: Questionnaire used by coaches for rating personal coaching implementation</i>	86
 Appendix C: Results of psychometric properties assessments	 93
<i>C1: Polychoric correlations for PBL items: first EFA</i>	93
<i>C2: Polychoric correlations for PBL items: repeated EFA</i>	94
<i>C3: Polychoric correlations for personal coaching items: EFA</i>	95
<i>C4: Overview of PBL structural equation models compared in CFA</i>	97
 Appendix D: Correlation scatter plots	 99
<i>D1. Scatter plots depicting session-level correlations for PBL fidelity scores</i>	99
<i>D2. Scatter plots depicting session-level correlations for personal coaching fidelity scores</i>	101

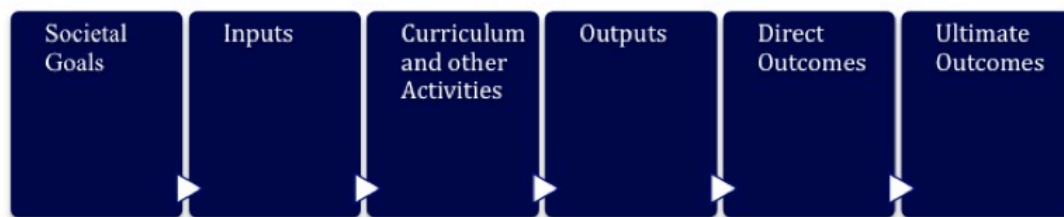
Appendix A: Delphi process materials and supplements

A1: Questionnaire used for Delphi Round One

Goal of this Panel Study: understanding the Studio Schools “Logic Model”

Dear colleague,

You were selected for this panel study because you have a good understanding of the Studio Schools model: either from having had a role in its design and development or from having worked with it. To inform our evaluation of Studio Schools, it is important to develop a detailed description of the programme’s components and how they relate to each other. We aim to do this through the development of a “*logic model*”, a graphical depiction of the logical relationships between the objectives, resources, activities, outputs and outcomes of a program. The goal of this panel study is to achieve consensus amongst a group of Studio School experts with regards to Studio Schools’ model and its components, and to describe them in such a way that a logic model that is usable for evaluative research can be developed. A graphical depiction of a logic model (tailored to the context of Studio Schools) is shown below.



Graphical depiction of a Logic Model

1. Societal Goals (Objectives)

Societal Goals normally refer to the purpose(s) of a programme in terms of addressing a given situation or problem in society as a whole (like high crime rates or a high proportion of people suffering from obesity). Please list what you think are the societal goals of Studio Schools.

Please list each societal goal in a separate textbox (as provided). For each societal goal, please clarify what you mean by it (if necessary) and why it is important in the separate explanation textbox (as provided). List a maximum of 8 societal goals. There is no lower limit of societal goals.

One societal goal of Studio Schools may be to decrease youth unemployment in England. An explanation could be a short elaboration on its relevance.

1. Please list Societal Goal 1 here:

Please explain what you mean by your description of Societal Goal 1 (if necessary) AND/OR why you believe it to be relevant here:

2. Please list Societal Goal 2 here:

Please explain what you mean by your description of Societal Goal 2 (if necessary) AND/OR why you believe it to be relevant here:

And so on until Societal Goal/Objective 8

2. Inputs

Inputs refer to the monetary or other resources a programme requires in order to function properly and be successful. Required inputs may be funds, staff or necessary materials. Please list what you think are the required inputs of a Studio School.

Please list each input in a separate textbox (as provided). For each input, please clarify what you mean by it (if necessary) and why it is important in the separate explanation textbox (as provided). List a maximum of 8 inputs. There is no lower limit of inputs.

One example of an Input of Studio Schools may be qualified teachers. An explanation could be your definition of a qualified teacher for a Studio School, and a short elaboration on his or her relevance.

1. Please list Input 1 here:

Please explain what you mean by your description of Input 1 (if necessary) AND/OR why you believe it to be relevant here:

2. Please list Input 2 here:

Please explain what you mean by your description of Input 2 (if necessary) AND/OR why you believe it to be relevant here:

And so on until Input 8

3. Ultimate (Distal) Outcomes

By ultimate outcomes, we mean the intended eventual impacts of a programme for those it caters for. In the case of Studio schools, these can include young people but if you wish these could also include employers or parents. Please list what you think are the ultimate outcomes of Studio Schools.

Please list each ultimate outcome in a separate textbox (as provided). For each ultimate outcome, please clarify what you mean by it (if necessary) and why it is important in the separate explanation textbox (as provided). List a maximum of 8 ultimate outcomes. There is no lower limit of ultimate outcomes.

One example of an ultimate outcome of Studio Schools for a student may be his or her increased employability. An explanation could be your definition of a student's employability and a short elaboration on its relevance.

An example of an ultimate outcome of Studio Schools for the wider community may be an enhanced workforce for local employers. An explanation could be a short elaboration on its relevance.

1. Please list Ultimate Outcome 1 here:

Please explain what you mean by your description of Ultimate Outcome 1 (if necessary) AND/OR why you believe it to be relevant here:

2. Please list Ultimate Outcome 2 here:

Please explain what you mean by your description of Ultimate Outcome 2 (if necessary) AND/OR why you believe it to be relevant here:

And so on until Ultimate (Distal) Outcome 8

4. Direct (Proximal) Outcomes

Direct outcomes are those outcomes that are a direct result of the efforts of and activities taking place as a part of a programme, and may in turn influence the ultimate outcomes. Please list what you think are the direct outcomes of studio schools.

Please list each direct outcome in a separate textbox (as provided). For each direct outcome, please clarify what you mean by it (if necessary) and why it is important in the separate explanation textbox (as provided). List a maximum of 8 direct outcomes. There is no lower limit of direct outcomes.

One example of a direct outcome of Studio Schools may be students' improved engagement with education (attendance, active participation). An explanation could be your description of students' improved engagement and a short elaboration on its relevance.

1. Please list Direct Outcome 1 here:

Please explain what you mean by your description of Direct Outcome 1 (if necessary) AND/OR why you believe it to be relevant here:

2. Please list Direct Outcome 2 here:

Please explain what you mean by your description of Direct Outcome 2 (if necessary) AND/OR why you believe it to be relevant here:

And so on until Direct (Proximal) Outcome 8

5. Activities: The Studio Schools Curriculum

As it is the objective of Studio Schools to teach the National Curriculum, we are not seeking your opinion on the subject content that is taught at Studio Schools. Rather, we are interested in the activities and methods by which Studio Schools deliver this curriculum. Please list what you think are the essential curricular activities taking place at a Studio School. Please note that we will ask you about non-curricular activities on the next page.

Please list each curricular activity in a separate textbox (as provided). For each curricular activity, please clarify what you mean by it (if necessary) and why it is important in the separate explanation textbox (as provided). List a maximum of 8 curricular activities. There is no lower limit of curricular activities.

An example of a curriculum component of Studio Schools may be Project-based Learning. An explanation could be your definition of Project-based Learning and a short elaboration on its relevance.

1. Please list Curricular Activity 1 here:

Please explain what you mean by your description of Curricular Activity 1 (if necessary) AND/OR why you believe it to be relevant here:

2. Please list Curricular Activity 2 here:

Please explain what you mean by your description of Curricular Activity 2 (if necessary) AND/OR why you believe it to be relevant here:

And so on until Curricular Activity 8

6. Activities: Non-curricular activities

By these we mean important activities other than those covered above under curriculum that you think are important for a Studio School to reach its objectives. Please list what you think are the important non-curricular activities taking place at Studio Schools.

Please list each non-curricular activity in a separate textbox (as provided). For each non-curricular activity, please clarify what you mean by it (if necessary) and why it is important in the separate explanation textbox (as provided). List a maximum of 8 non-curricular activities. There is no lower limit of non-curricular activities.

An example of such an activity may be the personal coaching of students. An explanation could be your description of the personal coaching process and a short elaboration on its relevance.

1. Please list Non-curricular Activity 1 here:

Please explain what you mean by your description of Non-curricular Activity 1 (if necessary) AND/OR why you believe it to be relevant here:

2. Please list Non-curricular Activity 2 here:

Please explain what you mean by your description of Non-curricular Activity 2 (if necessary) AND/OR why you believe it to be relevant here:

And so on until Non-curricular Activity 8

7. Outputs

Outputs refer to the concrete products of the activities that form part of a programme. Outputs can be directly linked to activities: an output of personal coaching can be a personalised curriculum for each student. Outputs should not be confused with outcomes (which are the effects outputs may have), or with activities (which generate outputs).

Please list what you think should be the outputs of the curricular and non-curricular activities of Studio Schools. Please list each output in a separate textbox (as provided). For each output, please clarify what you mean by it (if necessary) and why it is important in the separate explanation textbox (as provided). Also, please indicate which curriculum component/activity that you have specified under curriculum and activities (previous questions) should generate the output (a separate textbox is provided for this). List a maximum of 8 outputs. There is no lower limit of outputs.

1. Please list Output 1 here:

Please explain what you mean by your description of Output 1 (if necessary) AND/OR why you believe it to be relevant here:

Please list the curricular or non-curricular activity that generates Output 1 here:

2. Please list Output 2 here:

Please explain what you mean by your description of Output Activity 2 (if necessary) AND/OR why you believe it to be relevant here:

Please list the curricular or non-curricular activity that generates Output 2 here:

And so on until Output 8

8. Other relevant information

In case you think that there is any relevant information left regarding the components of the Studio Schools Model that you have not been able to share through this questionnaire, please indicate so below. If you have any other comments that you believe may be helpful to the researchers, please also share this information.

Additional Information:

Appendix A: Delphi process materials and supplements

A2: Original logic model items nominated by participants in Delphi Round One

Objectives

Participant 01 did not nominate items

Participant 02

1. Reduce youth unemployment
By better preparing young people for work and enabling them to thrive in the workplace
2. Contribute to economic growth
By developing a workforce more suited to the current labour market
3. Increase attainment
By providing a pedagogical approach more suited to certain types of students
4. Improve life outcomes
Due to strong correlation low non-cognitive skills and most negative life outcomes
5. Increase retention and engagement
By providing a curriculum more obviously relevant to life after school
6. Provide a genuine cross-over between vocational and academic education
By focusing on transferable employability skills
7. Increase wellbeing of young people
By focusing on personal coaching and the whole student

Participant 03

8. Increase transferable skills that young people develop during their time in school
9. Increase young people's awareness of the world of work through direct links with businesses
10. Decrease students becoming NEET in the future
11. Increase student engagement in routes of education
Evidence to indicate extent of disengagement by young people in school is high.

Participant 04

12. To improve the work readiness of young people in England

Participant 05

13. Decrease youth unemployment
Through work experience placements
14. More responsible citizens
Through being treated more like adults
15. Happiness
Personalisation provides students with direction to fulfil their potential, leading to a happier life in the short, medium and long-term
16. Employability / life skills
Through the CREATE framework
17. Health and well-being
Students are more aware of contemporary issues through PBL, e.g. road/traffic campaign
18. Corporate social responsibility
Employers engage with model through: placements, mentoring, guest speakers etc

Participant 06

19. To improve youth engagement in local communities
20. To develop communities of critical thinkers who are resilient and adaptable to change
21. To decrease youth unemployment by developing individuals with enterprising behaviour
22. To increase the number of students gaining valuable academic and vocational qualifications

Participant 07

23. Reduce youth unemployment in the School's catchment area
It is important for us to recruit young people in employment
24. Provide a legacy within the school's sector of specialty
Our specialism reflects our community and our goal is to provide a legacy within this: young people are reluctant to go to university due to the financial burden and so the Studio School can address the skill set for 14-19 yr olds
25. Develop skills for young people to be work ready by the age of graduation
Employers recognise a lack of skills beyond traditional academic qualifications and the Studio School aims to develop the acquisition of such skills

26. Community cohesion
The Studio School aims to develop and foster further links between the school and business, sector organizations, and wider community

Participant 08

27. To support emerging or prevalent sectors of the local economy.
28. To improve employability skills of young people
29. To make young people's educational experience more meaningful
Studio schools respond to young people's need to be treated as young adults and to be given real responsibility and autonomy. They reflect the fact that whereas working life has changed over the last few decades, most schools are still wedded to an industrial model of education that does not respond to young peoples expectations.
30. To offer more choice for young people so they can pursue their chosen field
31. To pioneer new ways of learning that better reflect industry, business and organisational activities
Project work is a key feature of Studio Schools and this is the vehicle for replicating how things happen in business and other organisations.
32. To promote choice and diversity in educational provision
This reflects our desire to have public services molded around our own needs and interests rather than having to adjust to what's on offer
33. To raise attainment by developing competition for existing schools
34. To respond to young people's need to do for meaningful work
There is a crisis of achievement particularly in white working class males that can be traced to the lack of well-paid employment opportunities. Studio schools bridge the aspiration gap by making real work and the prospect of it part of the educational experience. Its not something to look forward to; it something students are doing.

Participant 09

35. To increase participation in education
36. To provide an alternative route to employability that has rigour
37. To increase awareness of enterprise
38. To highlight the possibility of starting up a business

Participant 10

No objectives/societal goals were named

Inputs

Participant 01 did not nominate items

Participant 02

1. Students of all abilities for whom our pedagogical approach is most suitable
Studio Schools rely on attracting a mixed ability group
2. Teaching staff able to facilitate interdisciplinary project-based learning
This is outside most teachers' training or experience
3. Personal coaches able to support personal development and progression
4. Employers willing to provide work placements and other forms of engagement
5. A strong brand able to attract the most appropriate students and their parents plus employers
6. Supportive (or not hostile) local secondary schools from which students will be drawn
7. Curriculum and learning frameworks and materials support including related training
Interdisciplinary project based learning probably requires significantly more advance preparation from teachers
8. Sufficiently generous funding from DfE

Participant 03

9. Small schools
c 300 students in a school to enhance personalisation of provision for all young people
10. Personal Coaches
Professionals who would support in the development of the young people including both academic and work areas.
11. Project-based learning
Multi-disciplinary projects themed/based around the sector of employment the Studio School is based on, uniting different qualification outcomes.
12. Real work
Clear designated work for each young people with clear areas of responsibility.
13. CREATE framework
A mechanism for illustrating the development of non-cognitive skills.
14. Partnerships with businesses/sectors.

Participant 04

15. Sufficient revenue and capital funding from the DfE
16. Recruitment of a comprehensive intake of students
17. Support of local employers
18. National political support
19. Recruiting suitable staff

Participant 05

20. Coaches who are committed to PBL
21. Support and networking through the Studio Schools Trust
22. Suitable building
Space that has been designed for a studio school, e.g. space for project work; professional 'work-place' environment
23. Employers who provide work experience placements
24. Strong leadership and governance
Required to set up and run the school as well as manage the budget
25. Funding (DfE)
Capital funding; project development grant (pre-opening); diseconomy funding until school is at full-capacity; per-pupil funding
26. Projects (materials)
Curriculum mapped into projects
27. Equipment
e.g. specialist equipment to deliver specialism e.g. for creative and media

Participant 06

21. Teachers proficient in PBL
22. Visionary principals who can think creatively in the face of outside performance measures pressures
23. Monetary support of trust from partners to SST
24. Time to help the programme build up successfully
25. Local businesses to put in resource (time, money, personnel) to develop projects and work placements for young people
26. Effective personnel at SST to develop networking, resources, training, etc

Participant 07

27. Specialist staff to deliver curriculum and project based learning
28. Practitioners
Partners from the sector of specialism to share relevant skills and experiences in working practice within the sector industry
29. Personal coaches
Coaches to mentor the progression of all the students both academically and pastorally
30. Business manager
Partner to oversee, engage, foster and develop links with business
31. Specialist equipment
Materials and equipment to facilitate acquisition of relevant skills, techniques and methods of the sector industry
32. Partners
External partners to provide additional provision and support for young people
33. SENCO and teaching assistants
34. Visiting staff
Rolling programme of experts, practitioners, business partners, community members to engage, inspire and support young people

Participant 08

35. Qualified teachers and staff members with experience in the sector the studio school's specialist area
36. A working environment that reflects the best of the employment sector the students are focusing on
A Studio School should feel different to a normal school - it might import ways of working in the sector it is associated with
37. Committed partners from the sector and from progression routes eg universities
A distinguishing factor is the partnership base that feeds into the curriculum and experience of students through project setting, mentoring placements and employment opportunities
38. An appropriate building
A building that raises aspirations and promotes good relationships
39. Specialist equipment that is not available in a usual school
40. Furniture that creates an appropriate learning environment
This should be flexible to enable different types of learning engagement

- 41. ICT infrastructure that supports personalised learning
Studio schools are about independent learning so should enable this through software
- 42. Excellent support services that provide learning interventions for students
eg CAMHS, Family support workers, health workers, youth workers

Participant 09

- 43. Qualified teachers who are prepared to deliver a project based curriculum
- 44. Personal coaches
- 45. Local businesses
- 46. Inspirational speakers

Participant 10

- 47. Qualified teachers, to deliver high quality learning and teaching
- 48. Community and industry links.
- 49. Effective personal coaches.
- 50. Sufficient funding.
- 51. High quality vocational provision.
- 52. Support from Studio Schools trust.

Activities – Curriculum

Participant 01 did not nominate items

Participant 02

1. Interdisciplinary team project-based learning
2. Work placements, where those 16 and over are paid
3. Personal coaching sessions
4. Minimal single subject based learning

Participant 03

5. Enterprise projects
Project-based learning around themes connected to the host business and/or sector. Projects are aligned to a number of qualification outcomes and are cross-curricular.
6. Formal taught sessions
Some new concepts/ideas will be delivered in more traditional classroom/seminar settings so as to provide a starting point and focal direction for the following enterprise project.
7. Individual projects
Some work will be individual in nature to recognise the variety of projects and work in the real workplace.
8. Self-study units
Students will also require to complete some self-study units to bridge any gaps not covered by the enterprise projects and also to prepare for any examination subjects.
9. Portfolio
Students update a portfolio in relation to some qualification requirements but also to evidence and attribute their progression through the CREATE framework.

Participant 04

10. Project- and enquiry-based learning
11. CREATE skills
12. Work experience
13. Variable group sizes
14. Longer days at school and a longer school year

Participant 05

15. Project-based Learning
Curriculum is delivered through projects
16. Development of CREATE skills
Linked to projects and placements
17. Personalisation
Strong focus on personal development. Identifies where extra support is required; stretch activities for G&T;
18. Work experience
Linked to projects and CREATE
19. Small school / class sizes
Enable schools to deliver the model

Participant 06

20. Enterprise projects /PBL
21. Small group tuition
22. Work experience
23. Presentations of learning
24. Self / peer critique
25. Peer mentoring - particularly across the age range
26. Creating and running enterprises

Participant 07

27. Project based learning
Enquiry led projects will ensure that learners are practically researching, investigating and creatively presenting their work. Outcomes will be shared with the community and displayed to maximise the value of each project. Projects will cross over subjects and be paired to maximise the links in the learning; presentations of learning will allow student reflection on their learning journeys and enhance the opportunity for deep learning (meta cognition). Completed outcomes will be exhibited to develop a visual display of creative learning.
28. Contextual learning
Project based learning with make continual reference to the work related models applicable to a specialism. These replicate the final products of industry whilst creating the coursework of option A level and Btec courses. Students work across disciplines and learn about work briefs, collaboration, success and failure and how to learn from each of these experiences to gain resilience and insight into contemporary practice. Experience in a Studio

School is then reinforced by the work placements where acquired skills can be practiced in a real world setting

29. Curriculum enrichment

To enhance the core and option curriculum alongside the work placements, the school will run a programme of enrichment exploring the CREATE framework. Students will develop skills to explore the concept of citizenship and community through programmes including self directed projects building on their specialism and also with external programmes such as the National Citizenship Scheme

Participant 08

- 30. Lessons that use real life contexts drawn from the sector
- 31. Project Based Learning that is interdisciplinary and based in a real world context
- 32. A personal development programme that builds a sense of efficacy and agency
- 33. Coaching skills that are modelled by teachers and developed through peer coaching programmes
- 34. Professional courses and skills that equip students with experience of industry standard processes eg software
- 35. A personal coach who sets targets and shares accountability for progress
- 36. Self study time that promotes independent study skills
- 37. Work experience that is meaningful and develops real skills in the young person

Participant 09

- 38. Project based learning
- 39. Delivery of recognised qualifications through project based learning
- 40. CREATE
- 41. Personal coaching and learning coaching combined

Participant 10

- 42. Project based learning which is cross curricular.
- 43. Enterprise projects.
- 44. Extended work experience once a week for the duration of the course.
- 45. Explicit teaching of the CREATE skills framework.

46. Personal coaching which supports the student in skill development, personal development and careers.

Non-curricular activities

Participant 01 did not nominate items

Participant 02

1. Personal coaching of students
2. Opportunities to meet and engage with local employers
3. Opportunities to start an entrepreneurial project

Participant 03

4. Personal Coaches
Individual personalised support through an adult who is aware of all areas of development of the young person.
5. Formal relationship with business/business manager
Clear line management relationship within a business setting.
6. Real work (experience)
7. Involvement in the running of the Studio School
All students are involved in some capacity with the direction of the Studio School - whether via staff-student committee, or completing actual work to support with the operational side of the Studio School.
8. Real-world setting
Studio School should not feel like an annex within an educational facility; Studio Schools should 'feel' like a place of employment and work.
9. Longer working day and flexible working times (in some situations)
10. Students can have some flexibility in booking their holidays during the year.
11. Different structure to school year (to that of the traditional termly nature of schools).

Participant 04

12. National citizenship service
13. Community projects

Participant 05

14. Personal Coaching
15. Enrichment activities
e.g. sport / arts / special interest
16. Timetable and Calendar
The school day and year is longer than mainstream schools, to emulate the

working day/year and also enable the elements of the model to be delivered

17. Demand
There needs to be local demand for a SS from young people, parents and employers
18. Building and maintaining relationships with employers
To provide placements, projects etc
19. Effective leadership and governance
20. Active involvement with Studio Schools Trust and network of studio schools to share best practice, training etc

Participant 06

21. Personal coaching
22. Community volunteering
23. Extra curricular such as sports, arts related groups
24. Business Mentoring

Participant 07

25. Personal coaching of students
Personal coaches for each student will underpin the personalised learning that is key to a Studio School. Applicants find the idea of a mentor a very attractive prospect; students are looking for support in both academic and pastoral areas and the coach is seen to be the key provider in this. Coaches also work to support success in the work placements and so cover every area of Studio School life. They are key to attainment, retention, progression, engagement and happiness. As well as liaising with teaching staff, they liaise with the SENCO, business and parents to be at the heart of each student's learning experience.
26. Student led activities
Independence and self-motivation are to be nurtured to ensure an entrepreneurial spirit amongst all Studio School students. Staff will facilitate student led learning, enquiry and outcomes wherever and whenever possible. There are little social opportunities for the 14-19 age group here; student led projects will be key to success of the Studio School community. Projects can then be used for student entertainment and also to invite the wider community in.
27. Partnerships
Partnerships with other Studio Schools will be key to provide a sense of community between the students, reinforcing the values and beliefs of Studio School.

28. Touring events
To expand understanding of the sector, schools should host a programme of exhibitions, displays and workshops to ensure access to excellent sector provision. Visiting programmes to events etc should be offered to enrich the School experience.

Participant 08

29. A wide array of enrichment projects that challenge students to see themselves in unfamiliar situations.
30. Trips to workplaces to explore processes that are relevant to the curriculum and job roles
This is really curricular but often happens in non curriculum time
31. Fun events days that showcase new exciting aspects of the industry and allow students to be hands on
For example new product launches
32. Residential trips for teambuilding (induction), off site learning and university tasters
33. House competitions and inter school competitions
34. Sports competitions

Participant 09

35. Personal coaching
36. Business mentoring
37. Business coaching
38. Quality work experience

Participant 10

39. Personal coaching, although I don't really see this as non-curricular activity
40. Involvement in charity and community events to make students feel part of their wider community.
41. Extra curricular trips and visits e.g of a cultural or purely fun nature building a sense of cohesion.
42. Sports teams and fixtures with other schools to foster a team and school identity.

Outputs

Participant 01 did not nominate items

Participant 02

1. Personalised curriculum
2. Map of progression against the CREATE framework
3. Qualifications equivalent to minimum 5 A*-C GCSEs and 3 A Levels

Participant 03

4. Personalised curriculum
Output is generated by Enterprise projects and working with a Personal Coach
5. Individual targets (target setting)
Clear developmental points (in all aspects - both enterprise projects and the work areas)
Output is generated by meetings with a Personal Coach, and working directly with employers in the business.
6. Clear allocated work roles and responsibilities
Output is generated by real work and also working directly with employers in the business.
7. Completed academic qualifications
Output is generated by Portfolio/Self-study/Enterprise projects.
8. Development of non-cognitive skills by each student
Output is generated by enterprise projects/focus on the CREATE framework.
9. Increased engagement and satisfaction by young people in their education
Output is generated by personalised coach/enterprise projects/small school/set in a work environment.
10. Increase in positive relationships with adults
Output is generated by Small schools/personalised coaches/real work alongside employers

Participant 04

11. Examination outcomes
Output is generated by PBL and longer hours

Participant 05

12. Personalised curriculum
Output generated by Personal Coaching
13. Presentation of findings; curriculum covered;
Output generated by PBL

14. Employability skills
Output generated by Work Experience
15. Class sizes
Output generated by Small Schools
16. Extra-curricular qualifications/skills gained; membership of sports team;
Output generated by Enrichment
17. Number of quality of work placements
Output generated by Relationship with employers
18. Student recruitment
Output generated by Demand
19. Contribution to networking events; engagement with online network
Output generated by Membership of the Studio Schools Trust

Participant 06

20. Every pupil will gain work experience
Generated by work placements
21. Every pupil will get an opportunity to develop their CREATE employability skills
Generated by Coaching and PBL
22. Students will make links between curriculum areas
Generated by PBL
23. Students will have experience of presenting their learning
Generated by Presentations of Learning
24. Students will apply learning to the real world
Generated by PBL Enterprise Projects

Participant 07

25. Personalised curriculum
*Students have a personalised curriculum through support and discourse with their personal coach; this informs options, learning support, personal development, positive relationships, strategies for progression and regular points for review within the time at Studio School.
Output is generated by Personal Coaching*

Participant 08

26. Greater confidence and self-efficacy
Generated by personal coaching
27. Greater sense of responsibility
Generated by teamwork involved in Project Based Learning

- 28. Self awareness
Generated by Engaging with a personal development programme
- 29. Self-efficacy
Generated by following a project through from inception to celebration
- 30. Work-based skills
Generated by Work placement
- 31. Active engagement in own learning
Generated by Project Based Learning
- 32. Realistic assessment of career choices and progression routes
Generated by engagement with skills partners and mentors
- 33. Good attendance and punctuality
Generated by workplace engagement and Studio 'workplace' ethos

Participant 09

- 34. Personalised curriculum
- 35. Academic qualifications
- 36. Experience of the business world
- 37. High level employability skills

Participant 10

- 38. Personalised curriculum
Generated by Personal Coaching

Proximal Outcomes

Participant 01 did not nominate items

Participant 02

1. Higher attainment of students
2. Improved employment readiness
3. Higher retention and participation during school
4. Improved wellbeing, resilience
5. Better understanding of the world of work
6. Stronger portfolio to show employers

Participant 03

7. Students' improved engagement with education
8. Positive reception by community towards young people's roles in a Studio School.
Opportunity for the local citizens to observe/see/work with young people in a setting away from a traditional school.
9. Students have greater awareness of range and complexity of different skills for different circumstances (e.g. communication)
Students are able to evidence and indicate the nuances of such key skills as communication but are also able to deliver and evaluate such skills in a range of contexts such that their experience and understanding of it is both broader and deeper than in a traditional school.
10. Greater self-motivation amongst students
11. Greater awareness of self by students
Through greater focus on reflective and holistic elements
12. Increase in positive relationships between students and adults (both staff as well as people in the community)

Participant 04

13. Increased maturity of students
14. Improved attendance
15. Improved qualification outcomes

Participant 05

16. Improved attendance (students)
Students have chosen to be there. Learning by doing responds to students' learning style
17. Increased confidence (students)
Due to personalisation, pastoral support
18. Team-work skills (students)
As a result of PBL
19. More motivated students
Through 'learning by doing'
20. Improved attainment
21. More mature, professional students
Through development of CREATE skills
22. Improved behaviour

Participant 06

23. Students are more engaged in their learning
24. Teachers are passionate about the SS model and their profession
25. Value added is increased as a result of personalised learning and smaller class sizes
26. Students will have developed their CREATE skills
27. Students will be work ready as a result of work placements
28. Students will be self reflective and therefore apply themselves better in the workplace/in Higher Education
29. Improved literacy and numeracy skills of pupils as projects seek to explicitly develop these - making connections across subject areas.

Participant 07

30. Students' improved engagement with education
The project based delivery of the curriculum will enhance engagement with many learners; the personalised curriculum will also tailor make a provision to inspire and engage the learners
31. Students' attendance to improve
Engaging, project-based learning will inspire students and the practical methods will improve attendance and enjoyment of learning

- 32. Students stay in education
An engaging, interesting, specialised curriculum will encourage students to complete their education and enrol to continue until graduation at 18.
- 33. Improved community links
The curriculum is linked directly to the events in the area and with the whole community, students will be at the heart of the community to the mutual benefit of all concerned.

Participant 08

- 34. Students developing sector specific skills
- 35. Students gaining appropriate qualifications to progress to University courses, employment or entrepreneurship in the right area
- 36. Young people better able to meet workplace expectations such as commercial awareness and communication skills
- 37. Young people who are more self aware and therefore more able to make responsible choices about life and work
- 38. Young people who are confident because they have met expectations of the wider community in the presentation of their project work
- 39. Young people who are appreciative of a new way of learning that gives them responsibility
- 40. Young people who 'know what to do when they don't know what to do' - can think creatively and have learning skills
- 41. Young people who have the maturity and thus independence of mind to forge their own path and take advantage of emergent opportunities

Participant 09

- 42. Improved engagement with a diverse group of young people

Participant 10

- 43. Increased engagement with learning
- 44. Improved rates of attendance
- 45. Greater attainment in subjects students find more relevant to their interests and abilities.
- 46. Greater staying on rate in to further and higher education.

Distal Outcomes

Participant 01 did not nominate items

Participant 02

1. Higher rate of employment, employability and job satisfaction, lower drop out from work amongst alumni
Because students better able to identify, secure and retain the job they want
2. Higher economic growth
3. Higher attainment by students compared to what they would have achieved in a mainstream school
4. Higher level of personal development - confidence, self-awareness, resilience, wellbeing
5. Higher rate of entrepreneurial activity amongst alumni
6. Higher rate of participation and attainment in further or higher education of alumni

Participant 03

7. Increased employability (non-cognitive) skills
8. Increased employability
9. Increased/broader educational outcomes
10. Greater engagement by young people in their educational setting
11. Reduced NEETs
12. Secondary benefits that accompany reduction in NEETS: reduction in young people getting involved in crime, benefits, happiness etc..
13. Increased satisfaction/links between schools and employers

Participant 04

14. Enhanced earning power of Studio School alumni
15. Improvement in quality of local labour pool
16. Improved community cohesion

Participant 05

17. Increased employability for students
18. Enhanced workforce for local employers

19. Community cohesion
Through integrated projects and potential sharing of studio school building with the local community
20. Higher levels of maturity amongst young people
Greater exposure to adults and the working world
21. Aspirational young people
22. Greater choice of school for parents, carers and young people

Participant 06

23. More engaged adults in Britain who have a life long desire to learn and adapt
This will happen in many years to come
24. The British economy will be improved as enterprising behaviour will have been developed and more people will make a valuable contribution to society
25. Less crime in communities as more individuals take care of their communities
26. PBL will be proved to have produced well rounded individuals so other schools will borrow from SS model
27. Improved exam results as a result of increased student engagement in learning and attendance
28. Increased youth employment
29. All studio schools are judged 'outstanding'

Participant 07

30. Increased employability
Raised opportunities to attain employment within the sector industry or beyond; this could be full time employment or part time whilst students go to University
31. A legacy of the sector of specialism
The sector will have a new generation of practitioners to ensure the continuation of the industry in the UK
32. Local employment
It is hoped that graduates from a Studio School will create their own employment and businesses to support the regeneration of the area as well as create a bank of employees for established businesses.
33. Improved social cohesion
Partnerships between the young people, the school, business sector, and the local, national and international community will enhance social cohesion

- 34. Raised aspirations
Programmes of visiting staff, practitioners and experts along with displays, participation in large events and awareness of industry will raise aspirations of the young people at the Studio School.
- 35. Regeneration

Participant 08

- 36. Increased employability for young people
- 37. Breaking dependence on benefits for families
- 38. Greater community engagement of young people
Real world learning encourages the connection of young people to their communities and commitment to improve them
- 39. Increase in availability of skilled workers for the sector, and particularly for business partners
- 40. A more skilled workforce, more productive
- 41. Increase in global competitiveness in high skilled sectors and increased share of international contracts
- 42. Reduction in social alienation experienced by young people and thus a more stable society
- 43. Young people with wider skills from the CREATE framework and the self awareness focus of coaching who are able to improve their own quality of life and that of others

Participant 09

- 44. Increased employability skills
- 45. Young entrepreneurs who are able to look at establishing a successful business
- 46. Increase in local economy
- 47. Young people who have the skills to contribute to the local community and economy

Participant 10

- 48. Reduction in the number of students who are NEET.
- 49. Increased life chances for students who may have been failing in the education system prior to joining a studio school.
- 50. Greater employability of students leading to greater success at gaining and maintaining employment.

51. Employers able to recruit a suitably qualified and skilled workforce who are work ready on leaving school.
52. Greater confidence in students and better understanding of their own skills and abilities.

Appendix A: Delphi process materials and supplements

A3: List of non-redundant items, resulting from integration of nominated items by the researchers (rated by panellists in Delphi Round Two)

Approved by Panel prior to commencement of Round Two

Objectives

1. Poor work readiness, employability, and entrepreneurship of secondary school graduates
Employers recognise a lack of skills beyond traditional academic and cognitive skills that students have upon graduating from secondary school. These employability skills are important for students to be ready for work.
Nominated by rater 2
2. Poor educational attainment during secondary school, often leading to students dropping out
Many young people underperform in secondary school, with too many of them leaving secondary schools without an appropriate qualification.
Nominated by both raters
3. Poor engagement with education by young people during secondary school
Evidence suggests that the extent of disengagement (e.g., non-attendance, lack of interest, inactivity during class) by young people in school is high.
Nominated by rater 2
4. The large number of young people who are NEET in England
Nominated by both raters
5. The economic growth of local areas in England
Studio Schools can do this by helping to develop a workforce more suited to the current labour market
Nominated by both raters
6. The inadequacies of conventional secondary schools and a shortage of alternative options
Studio Schools provide a genuine cross-over between vocational and academic education that makes young people's educational experience more meaningful. This promotes choice and diversity in educational provision and reflects a desire to have public services molded around communities' needs.
Nominated by both raters
7. Poor community cohesion and young people's engagement with their communities
Studio Schools improve young people's responsibility by treating them more like adults, and aim to develop and foster links between the school and the wider community to improve youth engagement in local communities.
Nominated by both raters

Inputs

1. A sufficient number of appropriate teaching staff
Studio Schools require teaching staff who are able to facilitate interdisciplinary, project-based learning and who have experience in their school's specialist area. Other important staff may include a SENCO and sufficient teaching assistants.
Nominated by both raters
2. An appropriate and sufficient intake of students
Studio Schools rely on attracting a comprehensive body of students of mixed ability for whom this pedagogical approach is most suitable.
Nominated by both raters
3. Work placements offered by employers
Nominated by both raters
4. Personal Coaches
Personal coaches that are able to support the academic, vocational, and personal development of young people.
Nominated by both raters
5. Specialist materials and frameworks to facilitate the curriculum and learning
Studio Schools' distinctive approach requires specialist equipment (including an ICT infrastructure) that reflects a schools' sector of focus, is suitable for project-based learning, and supports personalised learning.
Nominated by both raters
6. Effective, visionary and unorthodox leadership
This is required to set up and run a school, manage its budget, and to think creatively in the face of external performance pressures.
Nominated by both raters
7. Funding
Studio Schools require sufficient funding from DfE and partners of the Studio Schools Trust (e.g., a pre-opening project development grant, diseconomy funding until the school is at full-capacity, and per-pupil funding)
Nominated by both raters
8. An appropriate building (physical space)
Studio Schools should be housed in a building that reflects a place of employment and work; they should raise aspirations, promote good relationships, and should 'feel' different to a normal school.
Nominated by both raters
9. Small school environment and class size
Studio Schools should have a maximum of (c.a.) 300 students in order to enable schools to deliver the model.
Nominated by both raters
10. High-quality, supportive relationships with civil and community partners
Studio Schools are new and require support (or at least no obstruction) from

organisations such as businesses/employers, partners from specialist sectors and industries, other secondary schools, and tertiary education institutions.

Nominated by both raters

11. Support and networking through the Studio Schools Trust
Active involvement with and support from the Studio Schools, including a schools' contribution to networking events and engagement with the online network.
Nominated by rater 1
12. Political and social support
Studio Schools are subject to local and national demand from young people, parents, and employers, and they require national political support (including time to get the model right).
Nominated by rater 2
13. A strong brand able to attract the most appropriate students and their parents plus employers.
Nominated by rater 1

Activities of Studio Schools

1. Project-based learning
Studio Schools should deliver the curriculum through multi-disciplinary projects that are based in real world contexts. Students work mostly in teams and occasionally work individually.
Nominated by both raters
2. Realistic work placements
Students should gain real work experience that is meaningful and develops necessary skills, where those over 16 and over are paid.
Nominated by both raters
3. Personal coaching sessions
Personal coaches should help students set targets and share accountability for students' progress. They support the student in skill development, personal development, and career development.
Nominated by both raters
4. Mentoring and coaching of students by local employers
Local employers should engage in professional relationships with Studio Schools students during which they provide guidance to the students.
Nominated by both raters
5. Opportunities to meet and engage with representatives of the professional world
Studio Schools students should get the opportunity to meet and engage with local employers and workplace through visits or trips, in order to explore processes that are relevant to the curriculum and job roles.
Nominated by rater 1

6. Some formal taught sessions, kept to a minimum
Some new concepts and ideas should be delivered in a more traditional classroom or seminar setting so as to provide a starting point and focal direction for a following enterprise project.
Nominated by both raters
7. Self-study units
To promote independent study skills and bridge any gaps not covered by PBL, and to prepare for any examination subjects, students should also require to complete some self-study units.
Nominated by both raters
8. Professional skills courses
Students should receive professional courses and skills that equip students with experience of industry standard processes.
Nominated by both raters
9. Enriching extra-curricular activities
These include school sports teams and intra/inter-school competitions, field trips, and events to foster a sense of cohesion and school identity.
Nominated by both raters
10. Involvement of students in the management of the Studio School
All students should be involved in some capacity with the direction of the Studio School – whether via a staff-student committee, or completing actual work to support with the operational side of the Studio School.
Nominated by both raters
11. Working with a unique school calendar
The structure of a Studio Schools school day and school year should be different to that of conventional schools, with a longer working day and school year, flexible working times in situations in which it is possible, and some flexibility for students in booking their holidays during the year.
Nominated by both raters
12. Community service undertaken by students
Students should engage in community projects and community volunteering - e.g., National Citizenship Service or involvement in charity- to make students feel part of their wider community.
Nominated by both raters
13. Entrepreneurial and self-initiated projects
Students should have the opportunity to start an entrepreneurial or enrichment project that challenges them to see themselves in unfamiliar situations.
Nominated by rater 2
14. Peer mentoring and coaching: particularly across the age range
Nominated by rater 1
15. Explicit teaching of the CREATE skills framework
Nominated by both raters

Outputs of the activities of Studio Schools

1. A personalised curriculum for all students
As a result of working with a personal coach and engaging in project-based learning, students should have a personalised learning plan. This includes a map of their progression against the CREATE framework, individual target setting, and strategies for personal development and academic progression.
Nominated by both raters
2. Students' portfolio
Students should update a portfolio in relation to some qualification requirements but also to evidence and attribute their progression through the CREATE framework and to be able to show employers their work experience.
Nominated by both raters
3. Students' work experience
Students should have gained work experience with clear allocated work roles and responsibilities.
Nominated by rater 2
4. Appropriate academic qualifications
Students should gain appropriate qualifications to progress to University courses, employment, or entrepreneurship in the right area, with a minimum of 5 A-C GCSE/s and 3 A-levels.*
Nominated by rater 2
5. Students' positive social networks
The activities of Studio Schools should foster students' positive social networks, including increases in positive relationships with adults in and beyond the schools and improved links with their communities.
Nominated by both raters
6. Students' extra-curricular achievements
Students should gain extra-curricular qualifications and skills, and an increase in memberships of civil associations (such as sports teams).
Nominated by rater 2

Proximal/Direct Outcomes of Studio Schools

1. Improved retention rates in school
The Studio Schools approach and curriculum encourages students to complete their secondary education.
Nominated by both raters
2. Students' improved engagement and satisfaction with education
Students are more engaged and satisfied with their education, resulting in higher attendance rates and active participation in class.
Nominated by rater 1

3. Students' increased educational attainment
Students' attainment in Studio Schools is higher than what they would have achieved in a conventional school, resulting in improved exam results and qualifications.
Nominated by both raters
4. Students' increased non-cognitive, CREATE skills
Students are able to evidence and indicate the nuances of these key skills.
Nominated by rater 1
5. Students' increased maturity and self-awareness
Students have the maturity and independence of mind to forge their own path, take advantage of emergent opportunities, and to think creatively to solve problems.
Nominated by rater 1
6. Students' improved confidence, aspirations, and self-efficacy
Because students have met expectations of the wider community through their work and projects, their confidence and self-efficacy is increased, which results in increased aspirations.
Nominated by rater 2
7. Students' improved health and well-being in school
Students' health and well-being is higher in a Studio School compared to what it would be like in a conventional school.
Nominated by both raters
8. Students' improved behaviour
Nominated by both raters
9. Teachers' increased job satisfaction
Teaching in a Studio School will increase teachers' passion about their profession and the model they work with.
Nominated by both raters
10. Students' improved relationships with their communities
Studio Schools' and their students' active engagement with local communities (e.g., through collaboration as part of projects, work experience, and extra-curricular activities) will lead to improved relationships with these communities for schools and students.
Nominated by both raters

Distal/Ultimate Outcomes of Studio Schools

1. Graduates' improved employability and work readiness
Students have a better understanding of the world of work and will be better able to meet workplace expectations (e.g., commercial awareness and communication skills) compared to what they would have achieved in a conventional school.
Nominated by rater 1

2. Higher employment rates and more success in employment for Studio Schools graduates
Students will be better able to identify, secure, and retain the job they want, and they will have a higher earning capacity compared to what they would have achieved in a conventional school. This results in a reduction in the number of young people who are NEET.
Nominated by rater 2
3. More (local) entrepreneurial activity by Studio Schools graduates
There will be a higher rate of entrepreneurial activity amongst alumni of Studio Schools compared to comparable alumni from conventional schools.
Nominated by both raters
4. Higher rates of enrolment and attainment in further or higher education amongst Studio Schools graduates
There will be a higher rate of participation and attainment in further or higher education of alumni compared to what it would have been like had they attended conventional schools.
Nominated by both raters
5. Graduates' improved long-term personal development
Students will have higher levels of important long-term development outcomes and skills such as confidence, self-awareness, resilience, and aspirations compared to when they would have attended conventional schools.
Nominated by rater 2
6. Graduates' improved life outcomes
Studio Schools students will have improved life outcomes in the long term such as happiness, health, and wellbeing compared to when they would have attended conventional schools.
Nominated by rater 1
7. Improved community welfare and engagement
Students will be more engaged in their local communities, leading to improved community cohesion, and they will less likely be NEET. As a result, there will be a reduction in social alienation experienced by young people and a reduction in negative outcomes, such as crime.
Nominated by both raters
8. An enhanced workforce for local employers or within the Studio Schools sector of specialism
Nominated by rater 1

Appendix A: Delphi process materials and supplements

A4: Results of Delphi Round Two

Objectives

- Poor work readiness, employability, and entrepreneurship of secondary school graduates

Employers recognise a lack of skills beyond traditional academic and cognitive skills that students have upon graduating from secondary school. These employability skills are important for students to be ready for work.

Participant	01	02	03	04	05	06	07	08	09	10
Rating	7	10	9	10	10	8	10	8	10	10
Summary	Median			Mode			Minimum		Maximum	
	10			10			7		10	
Comments - P 01: “A key prompt for the model and measure of whether its working”										

- Poor educational attainment during secondary school, often leading to students dropping out

Many young people underperform in secondary school, with too many of them leaving secondary schools without an appropriate qualification.

Participant	01	02	03	04	05	06	07	08	09	10
Rating	4	7	8	8	9	7	7	7	7	9
Summary	Median			Mode			Minimum		Maximum	
	7			7			4		9	
Comments - P 08: “Only if their curriculum has not met their needs in which case a studio school might meet them better”										

- Poor engagement with education by young people during secondary school
Evidence suggests that the extent of disengagement (e.g., non-attendance, lack of interest, inactivity during class) by young people in school is high.

Participant	01	02	03	04	05	06	07	08	09	10
Rating	8	7	10	8	10	9	9	5	10	10
Summary	Median			Mode			Minimum		Maximum	
	9			10			5		10	
Comments - P 01:“This is the key starting point - relevant to a significant minority of teenagers” - P 08 “I see it more as fulfilling niche interests that may or may not have led to lack of engagement; not as alternative provision”										

- The large number of young people who are NEET in England

Participant	01	02	03	04	05	06	07	08	09	10
Rating	5	5	8	7	9	6	5	5	6	10
Summary	Median			Mode			Minimum		Maximum	
	6			5			5		10	
Comments - P 01: “This clearly matters but the model should be useful for many young people who aren’t likely to become NEET” - P 08: “Not necessarily but you could see how the CREATE framework can underwrite personal goals and empowerment.”										

- The inadequacies of conventional secondary schools and a shortage of alternative options

Studio Schools provide a genuine cross-over between vocational and academic education that makes young people's educational experience more meaningful. This promotes choice and diversity in educational provision and reflects a desire to have public services molded around communities' needs.

Participant	01	02	03	04	05	06	07	08	09	10
Rating	5	8	9	8	8	9	8	9	10	10
Summary	Median			Mode			Minimum		Maximum	
	8.5			8			5		10	
Comments - P 01: “SS aim to provide an additional choice - not a universal option” - P 08: “Very much so - catering to niche interests and/or economic needs”										

- The economic growth of local areas in England

Studio Schools can do this by helping to develop a workforce more suited to the current labour market

Participant	01	02	03	04	05	06	07	08	09	10
Rating	4	9	8	7	9	5	8	8	10	9
Summary	Median			Mode			Minimum		Maximum	
	8			8			4		10	
Comments - P 08: “For sure; the schools respond to growth areas and fill skills gaps”										

- Poor community cohesion and young people's engagement with their communities

Studio Schools improve young people's responsibility by treating them more like adults, and aim to develop and foster links between the school and the wider community to improve youth engagement in local communities.

Participant	01	02	03	04	05	06	07	08	09	10
Rating	4	5	9	6	9	6	9	7	8	9
Summary	Median			Mode			Minimum		Maximum	
	7.5			9			4		9	
Comments - P 08: “Let’s hope so.”										

Inputs

- A sufficient number of appropriate teaching staff

Studio Schools require teaching staff who are able to facilitate interdisciplinary, project-based learning and who have experience in their school's specialist area. Other important staff may include a SENCO and sufficient teaching assistants.

Participant	01	02	03	04	05	06	07	08	09	10
Rating	9	9	10	9	10	10	10	10	10	10
Summary	Median			Mode			Minimum		Maximum	
	10			10			9		10	
Comments No comments were made										

- An appropriate and sufficient intake of students

Studio Schools rely on attracting a comprehensive body of students of mixed ability for whom this pedagogical approach is most suitable.

Participant	01	02	03	04	05	06	07	08	09	10
Rating	10	9	10	9	10	9	10	7	10	10
Summary	Median			Mode			Minimum		Maximum	
	10			10			7		10	
Comments No comments were made										

- Work placements offered by employers

Participant	01	02	03	04	05	06	07	08	09	10
Rating	9	8	10	9	10	8	10	4	8	10
Summary	Median			Mode			Minimum		Maximum	
	9			10			4		10	
Comments - P 08: “There are various ways of doing work placements; if the environment is a workplace or simulation then placement is not necessary”										

- Personal Coaches

Personal coaches that are able to support the academic, vocational, and personal development of young people.

Participant	01	02	03	04	05	06	07	08	09	10
Rating	9	7	10	10	10	8	10	9	10	9
Summary	Median			Mode			Minimum		Maximum	
	9.5			10			7		10	
Comments No comments were made										

- Specialist materials and frameworks to facilitate the curriculum and learning
Studio Schools' distinctive approach requires specialist equipment (including an ICT infrastructure) that reflects a schools' sector of focus, is suitable for project-based learning, and supports personalised learning.

Participant	01	02	03	04	05	06	07	08	09	10
Rating	9	5	10	8	10	9	9	10	10	8
Summary	Median			Mode			Minimum		Maximum	
	9			10			5		10	
Comments No comments were made										

- Effective, visionary and unorthodox leadership
This is required to set up and run a school, manage its budget, and to think creatively in the face of external performance pressures.

Participant	01	02	03	04	05	06	07	08	09	10
Rating	8	10	10	7	X	9	10	7	8	10
Summary	Median			Mode			Minimum		Maximum	
	9			10			7		10	
Comments - P 08: “There a limit to how unorthodox one can be but creativity will be important”										

- Funding
Studio Schools require sufficient funding from DfE and partners of the Studio Schools Trust (e.g., a pre-opening project development grant, diseconomy funding until the school is at full-capacity, and per-pupil funding)

Participant	01	02	03	04	05	06	07	08	09	10
Rating	10	6	10	8	10	10	10	10	10	10
Summary	Median			Mode			Minimum		Maximum	
	10			10			6		10	
Comments No comments were made										

- An appropriate building (physical space)
Studio Schools should be housed in a building that reflects a place of employment and work; they should raise aspirations, promote good relationships, and should 'feel' different to a normal school.

Participant	01	02	03	04	05	06	07	08	09	10
Rating	9	3	10	6	9	7	8	9	10	8
Summary	Median			Mode			Minimum		Maximum	
	8.5			9			3		10	
Comments No comments were made										

- Small school environment and class size

Studio Schools should have a maximum of approximately 300 students in order to enable schools to deliver the model.

Participant	01	02	03	04	05	06	07	08	09	10
Rating	7	9	10	8	10	10	9	2	7	8
Summary	Median			Mode			Minimum		Maximum	
	8.5			10			2		10	
Comments - P 08: “Small/intimate is good but 300 is arbitrary”										

- High-quality, supportive relationships with civil and community partners

Studio Schools are new and require support (or at least no obstruction) from organisations such as businesses/employers, partners from specialist sectors and industries, other secondary schools, and tertiary education institutions.

Participant	01	02	03	04	05	06	07	08	09	10
Rating	9	7	9	8	8	8	10	10	10	9
Summary	Median			Mode			Minimum		Maximum	
	9			8			7		10	
Comments										
- P 08: “Partners’ input into curriculum is key”										
- P 09: “A great deal of time and effort can be spent on overcoming this obstacle”										

- Support and networking through the Studio Schools Trust

Active involvement with and support from the Studio Schools, including a schools' contribution to networking events and engagement with the online network.

Participant	01	02	03	04	05	06	07	08	09	10
Rating	10	9	9	9	10	9	9	6	6	8
Summary	Median			Mode			Minimum		Maximum	
	9			9			6		10	
Comments - P 08: “Yes, although various studio models will flourish and there’s no single right model of student centred pedagogy”										

- Political and social support

Studio Schools are subject to local and national demand from young people, parents, and employers, and they require national political support (including time to get the model right).

Participant	01	02	03	04	05	06	07	08	09	10
Rating	9	4	9	8	9	10	10	9	10	10
Summary	Median			Mode			Minimum		Maximum	
	9			9			4		10	
Comments										
No comments were made										

- A strong brand able to attract the most appropriate students and their parents plus employers.

Participant	01	02	03	04	05	06	07	08	09	10
Rating	8	9	9	9	8	8	8	9	10	9
Summary	Median			Mode			Minimum		Maximum	
	9			9			8		10	
Comments - P 08: “Some studio schools seem to be very general in their offer so this would be less important”										

Activities

- Project-based learning

Studio Schools should deliver the curriculum through multi-disciplinary projects that are based in real world contexts. Students work mostly in teams and occasionally work individually.

Participant	01	02	03	04	05	06	07	08	09	10
Rating	10	10	10	8	10	10	10	7	9	9
Summary	Median			Mode			Minimum		Maximum	
	10			10			7		10	
Comments - P 10: “I disagree students work MOSTLY in team and only occasionally individually but feel projects are very important - Have scored on first point only and not on second.”										

- Realistic work placements

Students should gain real work experience that is meaningful and develops necessary skills, where those over 16 and over are paid.

Participant	01	02	03	04	05	06	07	08	09	10
Rating	10	8	10	8	10	7	9	7	8	10
Summary	Median			Mode			Minimum		Maximum	
	8.5			8			7		10	
Comments - P 06: “The paid element of this brought the mark down...the experience is more important than the payment.” - P 08: “Real world vocational learning yes, does this have to be in a work place setting or could they set up their own businesses? Or be freelancers?”										

- Personal coaching sessions

Personal coaches should help students set targets and share accountability for students' progress. They support the student in skill development, personal development, and career development.

Participant	01	02	03	04	05	06	07	08	09	10
Rating	9	8	10	9	10	9	10	9	10	9
Summary	Median			Mode			Minimum		Maximum	
	9			9			8		10	
Comments No comments were made										

- Mentoring and coaching of students by local employers

Local employers should engage in professional relationships with Studio Schools students during which they provide guidance to the students.

Participant	01	02	03	04	05	06	07	08	09	10
Rating	10	8	9	8	7	8	10	7	10	9
Summary	Median			Mode			Minimum		Maximum	
	8.5			8			7		10	
Comments - P 05: “Ideal, but not essential”										

- Opportunities to meet and engage with representatives of the professional world

Studio Schools students should get the opportunity to meet and engage with local employers and workplace through visits or trips, in order to explore processes that are relevant to the curriculum and job roles.

Participant	01	02	03	04	05	06	07	08	09	10
Rating	9	X	9	9	7	8	9	9	10	10
Summary	Median			Mode			Minimum		Maximum	
	9			9			7		10	
Comments - P 05: “Ideal, but not essential”										

- Some formal taught sessions, kept to a minimum

Some new concepts and ideas should be delivered in a more traditional classroom or seminar setting so as to provide a starting point and focal direction for a following enterprise project.

Participant	01	02	03	04	05	06	07	08	09	10
Rating	8	1	9	8	9	10	7	6	10	6
Summary	Median			Mode			Minimum		Maximum	
	8			8			1		10	
Comments - P 05: “Or, these sessions could be done after a project to fill in curriculum gaps” - P 10: “This suggests formal teaching is rare – which may not always be the case”										

- Self-study units

To promote independent study skills and bridge any gaps not covered by PBL, and to prepare for any examination subjects, students should also require completing some self-study units.

Participant	01	02	03	04	05	06	07	08	09	10
Rating	8	5	9	7	8	9	8	3	10	8
Summary	Median			Mode			Minimum		Maximum	
	8			8			3		10	
Comments - P 08: “I’m more concerned about students doing as well in a SS as they would at any school in term so of attainment PLUS having all the other stuff”										

- Professional skills courses

Students should receive professional courses and skills that equip students with experience of industry standard processes.

Participant	01	02	03	04	05	06	07	08	09	10
Rating	8	4	9	7	7	7	9	9	10	7
Summary	Median			Mode			Minimum		Maximum	
	7.5			7			4		10	
Comments - P 05: “Depends on the school”										

- Enriching extra-curricular activities

These can include school sports teams and intra/inter-school competitions, field trips, and events to foster a sense of cohesion and school identity.

Participant	01	02	03	04	05	06	07	08	09	10
Rating	8	7	10	8	10	9	10	8	6	8
Summary	Median			Mode			Minimum		Maximum	
	8			8			6		10	
Comments - P 10: “School sports team are difficult with small schools”										

- Involvement of students in the management of the Studio School

All students should be involved in some capacity with the direction of the Studio School – whether via a staff-student committee, or completing actual work to support with the operational side of the Studio School.

Participant	01	02	03	04	05	06	07	08	09	10
Rating	8	3	9	5	7	7	10	9	9	9
Summary	Median			Mode			Minimum		Maximum	
	8.5			9			3		10	
Comments - P 05: “Some students, not all”										

- Working with a unique school calendar

The structure of a Studio Schools school day and school year should be different to that of conventional schools, with a longer working day and school year, flexible working times in situations in which it is possible, and some flexibility for students in booking their holidays during the year.

Participant	01	02	03	04	05	06	07	08	09	10
Rating	8	7	9	9	10	7	10	2	X	4
Summary	Median			Mode			Minimum		Maximum	
	8			9			2		10	
Comments - P 06: “I do feel the working day should be more akin to the work place but the element of flexibility of holidays brought this rating down” - P 09: “Some flexibility but this really works better with 16 to 19 year olds. With 14 to 16 year olds we are restricted by the exam calendar”										

- Community service undertaken by students

Students should engage in community projects and community volunteering -e.g., National Citizenship Service or involvement in charity- to make students feel part of their wider community.

Participant	01	02	03	04	05	06	07	08	09	10
Rating	8	6	8	6	7	7	10	8	8	10
Summary	Median			Mode			Minimum		Maximum	
	8			8			6		10	
Comments - P 08: “Absolutely but this is not unique to studio schools”										

- Entrepreneurial and self-initiated projects

Students should have the opportunity to start an entrepreneurial or enrichment project that challenges them to see themselves in unfamiliar situations.

Participant	01	02	03	04	05	06	07	08	09	10
Rating	10	8	9	8	8	9	9	9	10	8
Summary	Median			Mode			Minimum		Maximum	
	9			8			8		10	
Comments - P 05: “This would suit some students”										

- Peer mentoring and coaching: particularly across the age range

Participant	01	02	03	04	05	06	07	08	09	10
Rating	9	9	9	7	8	7	7	7	6	5
Summary	Median			Mode			Minimum		Maximum	
	7			7			5		9	
Comments - P 05: “Good idea”										

- Explicit teaching of the CREATE skills framework

Participant	01	02	03	04	05	06	07	08	09	10
Rating	9	10	10	9	8	8	9	6	X	10
Summary	Median			Mode			Minimum		Maximum	
	9			9			6		10	
Comments - P 05: “CREATE isn't necessarily taught, rather students develop their CREATE skills” - P 08: “Implicit or explicit? Is this the right framework for all contexts? Should this evolve into a series of sector specific frameworks” - P 09: “CREATE should be built into all the learning and form the ethos of the SS”										

Outputs

- A personalised curriculum for all students

As a result of working with a personal coach and engaging in project-based learning, students should have a personalised learning plan. This includes a map of their progression against the CREATE framework, individual target setting, and strategies for personal development and academic progression.

Participant	01	02	03	04	05	06	07	08	09	10
Rating	10	10	10	7	10	9	10	9	10	5
Summary	Median			Mode			Minimum		Maximum	
	10			10			5		10	
Comments No comments were made										

- Students' portfolio

Students should update a portfolio in relation to some qualification requirements but also to evidence and attribute their progression through the CREATE framework and to be able to show employers their work experience.

Participant	01	02	03	04	05	06	07	08	09	10
Rating	10	9	10	7	7	10	10	8	7	7
Summary	Median			Mode			Minimum		Maximum	
	8.5			7			7		10	
Comments No comments were made										

- Students' work experience

Students should have gained work experience with clear allocated work roles and responsibilities.

Participant	01	02	03	04	05	06	07	08	09	10
Rating	10	8	10	8	8	9	10	7	X	9
Summary	Median			Mode			Minimum		Maximum	
	9			8			7		10	
Comments No comments were made										

- Appropriate academic qualifications

Students should gain appropriate qualifications to progress to University courses, employment, or entrepreneurship in the right area, with a minimum of 5 A-C GCSE/s and 3 A-levels.*

Participant	01	02	03	04	05	06	07	08	09	10
Rating	10	9	10	9	7	10	10	7	10	10
Summary	Median			Mode			Minimum		Maximum	
	10			10			7		10	
Comments - P 05: “Not necessarily A-Levels” - P 08: “To the best of their capability; not necessarily that benchmark”										

- Students' positive social networks

The activities of Studio Schools should foster students' positive social networks, including increases in positive relationships with adults in and beyond the schools and improved links with their communities.

Participant	01	02	03	04	05	06	07	08	09	10
Rating	10	8	10	7	8	8	10	9	10	9
Summary	Median			Mode			Minimum		Maximum	
	9			10			7		10	
Comments No comments were made										

- Students' extra-curricular achievements

Students should gain extra-curricular qualifications and skills, and an increase in memberships of civil associations (such as sports teams).

Participant	01	02	03	04	05	06	07	08	09	10
Rating	9	5	9	7	7	8	9	8	5	6
Summary	Median			Mode			Minimum		Maximum	
	7.5			9			5		9	
Comments No comments were made										

Proximal outcomes

- Improved retention rates in school

The Studio Schools approach and curriculum encourages students to complete their secondary education.

Participant	01	02	03	04	05	06	07	08	09	10
Rating	10	8	10	9	10	10	10	5	10	10
Summary	Median			Mode			Minimum		Maximum	
	10			10			5		10	
Comments - P 08: “In some studios”										

- Students’ improved engagement and satisfaction with education

Students are more engaged and satisfied with their education, resulting in higher attendance rates and active participation in class.

Participant	01	02	03	04	05	06	07	08	09	10
Rating	10	X	10	8	10	10	10	9	10	10
Summary	Median			Mode			Minimum		Maximum	
	10			10			8		10	
Comments No comments were made										

- Students’ increased educational attainment

Students’ attainment in Studio Schools is higher than what they would have achieved in a conventional school, resulting in improved exam results and qualifications.

Participant	01	02	03	04	05	06	07	08	09	10
Rating	10	10	9	8	10	10	10	9	10	10
Summary	Median			Mode			Minimum		Maximum	
	10			10			8		10	
Comments No comments were made										

- Students’ increased non-cognitive, CREATE skills

Students are able to evidence and indicate the nuances of these key skills.

Participant	01	02	03	04	05	06	07	08	09	10
Rating	10	10	9	9	10	10	10	9	9	7
Summary	Median			Mode			Minimum		Maximum	
	9.5			10			7		10	
Comments No comments were made										

- Students' increased maturity and self-awareness

Students have the maturity and independence of mind to forge their own path, take advantage of emergent opportunities, and to think creatively to solve problems.

Participant	01	02	03	04	05	06	07	08	09	10
Rating	10	9	9	8	10	10	10	9	10	10
Summary	Median			Mode			Minimum		Maximum	
	10			10			8		10	
Comments No comments were made										

- Students' improved confidence, aspirations, and self-efficacy

Because students have met expectations of the wider community through their work and projects, their confidence and self-efficacy is increased, which results in increased aspirations.

Participant	01	02	03	04	05	06	07	08	09	10
Rating	10	9	9	8	8	9	10	9	10	9
Summary	Median			Mode			Minimum		Maximum	
	9			9			8		10	
Comments										
No comments were made										

- Students' improved health and well-being in school

Students' health and well-being is higher in a Studio School compared to what it would be like in a conventional school.

Participant	01	02	03	04	05	06	07	08	09	10
Rating	9	4	9	8	8	8	9	4	8	10
Summary	Median			Mode			Minimum		Maximum	
	8			8			4		10	
Comments										
- P 08: “Depends on how much focus this is given”										

- Students' improved behaviour

Participant	01	02	03	04	05	06	07	08	09	10
Rating	9	5	9	8	9	9	10	X	10	8
Summary	Median			Mode			Minimum		Maximum	
	9			9			5		10	
Comments										
- P 08: “More autonomy should lead to better behaviour but there could be a lag”										

- Teachers' increased job satisfaction

Teaching in a Studio School will increase teachers' passion about their profession and the model they work with.

Participant	01	02	03	04	05	06	07	08	09	10
Rating	9	5	9	7	8	9	9	8	10	8
Summary	Median			Mode			Minimum		Maximum	
	8.5			9			5		10	
Comments No comments were made										

- Students' improved relationships with their communities

Studio Schools' and their students' active engagement with local communities (e.g., through collaboration as part of projects, work experience, and extra-curricular activities) will lead to improved relationships with these communities for schools and students.

Participant	01	02	03	04	05	06	07	08	09	10
Rating	9	7	10	X	9	8	10	9	8	8
Summary	Median			Mode			Minimum		Maximum	
	9			8			7		10	
Comments No comments were made										

Distal outcomes

- Graduates' improved employability and work readiness

Students have a better understanding of the world of work and will be better able to meet workplace expectations (e.g., commercial awareness and communication skills) compared to what they would have achieved in a conventional school.

Participant	01	02	03	04	05	06	07	08	09	10
Rating	10	9	10	9	10	10	10	9	10	7
Summary	Median			Mode			Minimum		Maximum	
	10			10			7		10	
Comments No comments were made										

- Higher employment rates and more success in employment for Studio School graduates

Students will be better able to identify, secure, and retain the job they want, and they will have a higher earning capacity compared to what they would have achieved in a conventional school. This results in a reduction in the number of young people who are NEET.

Participant	01	02	03	04	05	06	07	08	09	10
Rating	10	10	8	9	9	10	10	7	9	9
Summary	Median			Mode			Minimum		Maximum	
	9			9			7		10	
Comments										
No comments were made										

- More (local) entrepreneurial activity by Studio School graduates

There will be a higher rate of entrepreneurial activity amongst alumni of Studio Schools compared to comparable alumni from conventional schools.

Participant	01	02	03	04	05	06	07	08	09	10
Rating	10	5	9	7	9	7	10	8	10	8
Summary	Median			Mode			Minimum		Maximum	
	8.5			10			5		10	
Comments No comments were made										

- Higher rates of enrolment and attainment in further or higher education amongst Studio School graduates

There will be a higher rate of participation and attainment in further or higher education of alumni compared to what it would have been like had they attended conventional schools.

Participant	01	02	03	04	05	06	07	08	09	10
Rating	9	7	9	7	7	9	10	7	8	8
Summary	Median			Mode			Minimum		Maximum	
	8			7			7		10	
Comments										
No comments were made										

- Graduates' improved long-term personal development

Students will have higher levels of important long-term development outcomes and skills such as confidence, self-awareness, resilience, and aspirations compared to when they would have attended conventional schools.

Participant	01	02	03	04	05	06	07	08	09	10
Rating	10	9	10	9	9	10	10	9	10	8
Summary	Median			Mode			Minimum		Maximum	
	9.5			10			8		10	
Comments No comments were made										

- Graduates' improved life outcomes

Studio School students will have improved life outcomes in the long term such as happiness, health, and wellbeing compared to when they would have attended conventional schools.

Participant	01	02	03	04	05	06	07	08	09	10
Rating	10	10	9	8	8	9	10	8	8	9
Summary	Median			Mode			Minimum		Maximum	
	9			8			8		10	
Comments										
No comments were made										

- Improved community welfare and engagement

Students will be more engaged in their local communities, leading to improved community cohesion, and they will less likely be NEET. As a result, there will be a reduction in social alienation experienced by young people and a reduction in negative outcomes, such as crime.

Participant	01	02	03	04	05	06	07	08	09	10
Rating	9	3	9	8	9	8	10	5	8	8
Summary	Median			Mode			Minimum		Maximum	
	8			8			3		10	
Comments - P 08: “If the studio model is successful and adopted by mainstream schools; then would it still be a studio?”										

- An enhanced workforce for local employers or within the Studio Schools sector of specialism

Participant	01	02	03	04	05	06	07	08	09	10
Rating	10	3	9	8	9	7	10	9	10	8
Summary	Median			Mode			Minimum		Maximum	
	9			9			3		10	
Comments No comments were made										

Appendix A: Delphi process materials and supplements

A5: Decision process on items classified as uncertain after round two of the Delphi

During this phase, items that were classified as uncertain based on the panel's response during the second round were then included or excluded by the first author. For each uncertain item, the decision to include or exclude was guided by the objective to develop a logically coherent set of items that describes the defining features of Studio Schools. Decisions were made based on information obtained from the following sources:

*Numerical (ratings) and written responses provided by participants in the Rounds One and Two questionnaires.

*Semi-structured interviews that were conducted with some members of the panel prior to the Round One questionnaire.

*Documentation provided by the Studio Schools Trust and found on the Studio Schools online network.

For each decision, the source from which the information to guide the decision was obtained is referred to. Exhibit 2 shows the items that were included and excluded by the researcher.

Exhibit A5.1: items included and excluded by the first author

Included items

Societal goals

- The economic growth of local areas in England

Reason: the Studio Schools research report (p. 5) explicitly cites economic problems as part of the need for Studio Schools. Though the dispersion of the panel's quantitative responses led to an initial rating of "uncertain," the importance of including this concept reflects in the central tendencies of the panel's quantitative response (median=8, mode=8)

- Poor educational attainment during secondary school, often leading to students dropping out

Reason: although less emphasized than engagement, poor attainment in education is explicitly cited in the Studio Schools research report (p. 3) as part of the need for Studio Schools. This reflects in the panel's quantitative response (median=7, mode=7)

Inputs

- Effective, visionary and unorthodox leadership: changed to 'Effective, visionary and creative leadership'

Reason: this item was highly rated by the panel (median=9, mode=10), with one of the two panel members who rated the item with a value lower than 8 stating: 'there is a limit to how unorthodox one can be but creativity will be important'. Thus, the item was rephrased and included in the logic model.

- Small school environment and class size

Reason: this is mentioned as a core feature of the model throughout documents and interviews, and is one of the 7 essential elements of a Studio School (Introduction to the Studio Schools model, p. 5)

Activities

- Realistic work placements

Reason: this is mentioned as a core feature of the model throughout documents and interviews, and is one of the 7 essential elements of a Studio School (Introduction to the Studio Schools model, p. 5).

- Mentoring and coaching of students by local employers

Reason: this received high ratings (median=8.50, mode=8) and can be integrated with the realistic work placements that form an essential part of the model

- Some formal taught sessions, kept to a minimum

Reason: calendars and timetables provided by the SST and schools on the network suggest that this is often done. This is supported by the quantitative (median=8, mode=8) and qualitative ('this suggests formal teaching is rare – this may not always be the case') response by the panel. It seems that this is one of the flexible elements of the model, so the 'kept to a minimum' part is removed from the logic model.

- Self-study units

Reason: independent learning units are a flexible option according to the Timetabling Overview provided by the SST on the network. The potential importance of this item is reflected in the panel's quantitative response (median=8, mode=8) although there is no firm agreement. This item is therefore integrated with the other school-based activities

- Working with a unique school calendar

Reason: although the dispersion in ratings by the panel suggests that this may vary somewhat from school to school, the Timetabling Overview provided by the SST recommends a longer school year with longer school days, and to some extent flexibility in booking holidays. The potential importance of this item is underpinned by the (despite the dispersion) relatively high ratings by the panel (median=8, mode=9).

Outputs

- A personalised curriculum for all students

Reason: This is one of the 7 essential elements of a Studio School (Introduction to the Studio Schools model, p. 5). This also reflects in overall high ratings by the panel (median=10, mode=10).

- Appropriate academic qualifications

Students should gain appropriate qualifications to progress to University courses, employment, or entrepreneurship in the right area, with a minimum of 5 A*-C GCSE/s and 3 A-levels

Reason: the panel members who rated this item under 9 expressed concerns about the description of this item. The introduction to the Studio Schools model provided by the trust (p. 6) states that all schools will deliver qualifications at Level 2 and above. Thus, this item was included and the final part (5 GCSEs and 3 A-levels) was rephrased to 'with a minimum of a Level 2 equivalent'.

Direct outcomes

- Students' improved health and well-being in school

Reason: students' improved wellbeing is cited in interviews as an important result the features of the model should have, which reflects in the panel's overall rating of this item (median=8, mode=8) The benefits of small schools described in the SST's Introduction to the Studio Schools model (p.8) also contribute to students' wellbeing. No support was found for the health element, which was therefore dropped.

Ultimate outcomes

- An enhanced workforce for local employers or within the Studio Schools' sector of specialism

Reason: the concerns raised by employers about the employability of school leavers were one of the original drivers for Studio Schools (see Introduction to the Studio Schools model, p. 8). This reflects in the high overall ratings by the panel for this item (median=9, mode=9)

- More (local) entrepreneurial activity by Studio Schools graduates

Reason: although there is no explicit reference to entrepreneurial activity in the documents and interviews, the overall rating for this item by the panel was very high (median= 8,5, mode= 10), and this reflects employability (not all employable students join business or organizations, some found their own). Therefore, this item was integrated with graduates' improved employability.

- Improved community welfare and engagement

Reason: the involvement and active engagement of employers, community organisations, and the (local) community in general is cited throughout documents and interviews as an important element of the Studio Schools approach. This also reflects in the overall ratings by the panel (median=8, mode=8)

Excluded items

Societal goals (rephrased as objectives)

- The large number of young people who are NEET in England

Reason: from the panel's quantitative (median=6, mode=5) and qualitative ('This clearly matters but the model should be useful for many young people who aren't likely to become NEET') response, it became clear that although this may be a welcome result of Studio Schools, it actually is not one of the objectives of the model.

- Poor community cohesion and young people's engagement with their communities

Reason: despite relatively high ratings by the panel (median=7.5, mode=9), the need to improve the cohesion of wider communities and the engagement of young people with their communities is not mentioned as an objective in documentation and during interviews with respondents. No reasons were reported in support of high ratings by the panel. Thus, although Studio Schools may improve community cohesion and students' engagement with their communities, this is not recognised as an actual objective of the model.

Inputs

- An appropriate building (physical space)

Reason: despite relatively high overall ratings by the panel (median=8.5, mode=9), the need for an appropriate building (that reflects a workplace, raises aspirations, promotes good relationships and feels different to a normal school) is not mentioned in documentation and during interviews with respondents. Thus, this is not included as a necessary input for success.

Activities

- Enriching extra-curricular activities

Reason: there is no explicit support for the need for these activities for the model to function properly provided in documentation, interviews, or by the panel during in the questionnaires.

- Community service undertaken by students (ratings ranged from 6-10)

Reason: The SST National Citizen Service Pilot 2013 suggests that this is (at this point) an optional feature, rather than a defining feature of Studio Schools. This is understated by qualitative response from the panel: 'absolutely but this is not unique to studio schools'.

- Professional skills courses

Reason: there is no explicit support for the need for professional skills courses for the model to function properly provided in documentation, interviews, or by the panel during the questionnaires. Although this may be important to students, qualitative response from the panel suggests that this 'depends on the school'. Thus, this was excluded from the logic model that is relevant to all schools.

- Involvement of students in the management of the Studio School

Reason: although this may be an important part of policies in some schools, there is no explicit support provided for the need for this in documents, interviews or by the panel. Qualitative response suggests that this may be relevant for 'some students, not all'. This appears to be an optional, but not a defining feature of a Studio School.

- Peer mentoring and coaching: particularly across the age range

Reason: although this may be a valuable addition to the model, there is no explicit support provided for this item in the documentation, interviews, or by the panel. It appears as if this is not a defining feature of Studio Schools. Thus, it was excluded from the logic model.

Outputs

- Students' portfolio Students should update a portfolio in relation to some qualification requirements but also to evidence and attribute their progression through the CREATE framework and to be able to show employers their work experience

Reason: although, despite dispersion, this item received high ratings from the panel (median=8,5, modes are 7 and 10), there is no support provided for the need for students to keep an explicit portfolio in documents, interviews or the panel response. This suggest that there are other ways through which students might progress through the CREATE framework and show their work experience.

- Students' extra-curricular achievements

There is no explicit support provided for the explicit need for students to keep a portfolio in documents, interviews or the panel response. This appears to be a welcome addition to the Studio Schools model, but not one of its defining features.

Direct outcomes

- Teachers' increased job satisfaction

Reason: although this may be a welcome result of the Studio Schools model, students form the target population of the model. Therefore, this is not considered a defining feature of the model.

Ultimate outcomes

- Higher rates of enrolment and attainment in further or higher education amongst Studio Schools graduates

Reason: although this is a welcome result of any secondary education experience, there is no explicit support in the documents, interviews or panel response for this item as a defining feature of the Studio Schools model.

Appendix B: Implementation checklists and questionnaires

B1: Checklist used by researcher for rating PBL implementation

1. Please rate the accuracy of these statements based on your observations of the current project session

How much do you agree with the following statements?

Question	Answer			
	Completely accurate	Somewhat accurate	Somewhat inaccurate	Completely inaccurate
a. The people involved in this project should inspire students to put in their best effort				
b. Students had sufficient opportunities to share ideas about subject content to others during this session				
c. During this session, students have used key methods of enquiry (e.g. document analysis for English/History, conducting experiments for Science)				
d. During this session, students spent most of their time working on their projects				
e. Students' application of knowledge and skills within the context of this project reflects how these are applied in the real world				
f. During today's session, students have had to reflect explicitly on which of their CREATE-skills they need to further develop to successfully complete the project				

How often do you think the following statements are accurate?

Question	Answer			
	Always	Most of the time	Most often not	Never
g. When the teacher assisted students with their project work during this session, she/he offered solutions to the problems they encountered				
h. During this session, the teacher instructed students about what they need to study to complete the project successfully				
i. While working on this project during this session, students discovered independently where to find the information they need to complete the project				
j. During this session, all students have received feedback feedback/critique on their project work from their learning coach				
k. During this session, students have provided/received feedback on the products of others'/their project activities				

Do you agree with the following statements?

Question	Answer		
	Yes	No	Not sure
l. This project was presented to students with a clear explanation of the steps they needed to finish to complete their assignment			
m. During this session, students have worked or met with at least one adult from outside the school with relevant expertise/experience			

2. Please rate the accuracy of these statements based on the qualitative data obtained from teachers and students regarding the current session

Question	Answer			
	Completely accurate	Somewhat accurate	Somewhat inaccurate	Completely inaccurate
a. It is clear to teacher how the subject matter studied by students aligns with what they need to know to complete this project successfully				
b. It is clear to students how the subject matter studied by students aligns with what they need to know to complete this project successfully				
c. It is clear to teacher how the content/knowledge that students study to complete their project links to the content/knowledge they need to learn to pass their exams				
d. It is clear to students how the content/knowledge that students study to complete their project links to the content/knowledge they need to learn to pass their exams				

Base this on answers to question 4a of the weekly checklists.

Question	Answer			
	Completely accurate	Somewhat accurate	Somewhat inaccurate	Completely inaccurate
e. During today's session, students have used content from one or more subjects that are offered at this School (i.e. English language, English literature, math, science) to complete project tasks				

Base this on answers to question 4b of the weekly checklists.

Question	Answer			
	Completely accurate	Somewhat accurate	Somewhat inaccurate	Completely inaccurate
f. It is clear to teacher how the knowledge and skills students obtain in the context of this project work can have utility in other contexts				
g. It is clear to students how the knowledge and skills students obtain in the context of this project work can have utility in other contexts				

Base this on answers to question 4c of the weekly checklists.

Appendix B: Implementation checklists and questionnaires

B2: Questionnaire used by students for rating PBL implementation

INTRODUCTION

- ☆ **We are researchers and we would like to get your views about the project-lesson you have just taken. This is NOT a test. There are no right or wrong answers!**
- ☆ **Your name is not on this questionnaire. This questionnaire is COMPLETELY CONFIDENTIAL. No-one will find out what you write other than the researchers. Your family, friends and school will NOT find out what you write.**
- ☆ **The goal of this questionnaire is NOT to see how well you are doing in school or how well your learning coach does her or his job. We just want to see how your school is working with this important part of the Studio School Model. No-one other than the researcher will find out what you write about yourself or your learning coach.**

COMPLETING THE QUESTIONNAIRE

- ☆ **Please answer all the questions as truthfully as you can. Please fill it all in and be honest. The answers you give will be used to learn more about how Studio Schools work.**
- ☆ **Please put a tick (✓) in the boxes provided. If you have made a mistake, put a cross (x) through the box and answer again.**

Thanks very much for your help!

Before you complete this questionnaire, please fill in your participant ID (given to you by the researcher) and today's date.

PARTICIPANT ID	
DATE	

1. These questions are about THE PROJECT LESSON THAT YOU HAVE JUST FINISHED.

HOW MUCH DO YOU AGREE with the following statements?

Question	Answer			
	Totally Agree	Agree a bit	Don't really agree	Totally disagree
a. The people involved in this project inspire me to try my best				
b. I DON'T feel that I had enough chances to share my ideas about what I learned with others during this lesson				
c. It is clear to me which CREATE-skills I need to develop further to do this project well				
d. During this lesson, my classmates and I used useful research skills (e.g. studying texts for English, doing experiments for Science)				
e. I am NOT sure how the things I learn as part of this project can be useful in other situations				
f. Today, I have had to think about which of my CREATE skills I need to develop further to do this project well				
g. During this lesson, my classmates and I spent most of our time doing project work				
h. I DON'T think the way I use my skills and knowledge in this project is the same as the ways these are used in the real world				
i. I feel that this project belongs to my classmates and me, and that it is our job to complete it well				
j. I feel that I am developing my own best way of studying so that I can put what I learn to good use for project tasks				
k. It is clear to me how the project tasks I need to complete help me develop specific CREATE-skills				
l. I DON'T understand how the things I study to complete my project link to what I need to for my exams				
m. It is clear to me how the things I study (e.g. English, Science, or Maths) are important to know in order to complete this project well				
n. I recognise aspects of my own work that can or should be improved				
o. The problems we solve with this project are important to me and my classmates				

p. I know the aspects of my own work that are particularly strong				
---	--	--	--	--

2. These questions are also about THE PROJECT LESSON THAT YOU HAVE JUST FINISHED.

HOW OFTEN do you think the following statements are true?

Question	Answer			
	Always	Most of the time	Most often not	Never
a. When our learning coach helped us with our project work, she/he offered solutions to the problems we came across				
b. To complete my project tasks today, I have had to use knowledge from one or more core subjects (e.g. English or Science)				
c. During this lesson, my classmates and I found out for ourselves where to find the information we need to complete the project				
d. Today, I gave feedback to my classmates on the assignments that we did as part of this project and also and received it from them on my own work				
e. Today, my learning coach gave me feedback on the assignments I did as part of this project				
f. During this lesson, the learning coach told us what we need to learn to complete the project well				

3. These questions are also about THE PROJECT LESSON THAT YOU HAVE JUST FINISHED.

DO YOU AGREE with the following statements?

Question	Answer		
	Yes	No	Don't know
a. This project was presented to us with a clear explanation of the steps we needed to finish to complete our assignment			
b. Today, I worked or met with at least one adult from outside the school who is an expert on the project topic when I was completing my project work			

4. These questions are about the things you have done and learned in class today. Please answer these questions honestly and as completely as possible.

a. Please describe what you have studied or learned today for this project and how you think it is:

I) Important to complete this project successfully

II) Important for what you need to know for your exams

Information studied or learned: (e.g. part of a textbook or something you found on the internet. If you don't feel that you have learned anything today, then that's an answer too)	This is important to complete the project because:	This is important for my exams because:

b. Please describe:

I) The subject or subjects you have had to use today while working on the project.

II) What knowledge you have used from each subject.

Subject: (e.g. English or Maths)	Used knowledge: (e.g., writing, running analyses, creating charts):

c. Please describe how you think the things you have learned for this project today can be useful in other situations. (e.g. at work or at home)

Information, Knowledge or Skills learned today as part of this project: (e.g. You may have learned more English, Math, or Science knowledge or you may have learned more communicating skills)	How it can be useful in other situations:
--	---

Appendix B: Implementation checklists and questionnaires

B3: Questionnaire used by teachers for rating PBL implementation

INTRODUCTION

- The University of Oxford and the Studio Schools Trust (SST) are working on a preliminary evaluation of the Studio Schools model. As part of this, we would like to ask you to complete a questionnaire regarding the project session that you have just thought.
- This questionnaire is completely confidential. No party other than the researchers will see your answers. Your school and the Studio Schools Trust will not see your answers.
- It is NOT the purpose of this questionnaire to monitor your individual performance or to prescribe how to teach a session of project-based learning. Merely, the questionnaire seeks to gain insight into how your school works with this important aspect of the Studio School model. No other party than the researchers will see your individual answers or the answers your students give about your specific project sessions.

COMPLETING THE QUESTIONNAIRE

- Please complete all questions as truthfully as possible. Your answers will be used to learn more about the ways Studio Schools work with the Studio Schools model and the factors that may influence how they do so.
- Please put a tick (✓) in the boxes provided. In case you would like to rectify any mistakes, please put a cross (x) through the box and answer again.

Thank you very much for your help!

Before you complete this questionnaire, please fill in your participant ID (given to you by the researcher) and today's date.

PARTICIPANT ID	
DATE	

1. The following questions are about **THE PROJECT SESSION THAT YOU HAVE JUST TAUGHT.**

How much do you agree with the following statements?

Question	Answer			
	Completely agree	Somewhat agree	Somewhat disagree	Completely disagree
a. The people involved in this project inspire students to put in their best effort				
b. I do not think that students had sufficient opportunities to share ideas about subject content to others during this session				
c. It is clear to me which CREATE-skills each of my students need to develop further to complete this project successfully				
d. During this session, students have used key methods of enquiry (e.g. document analysis for English/History, conducting experiments for Science)				
e. I am not sure about how the knowledge and skills students obtain in the context of their project work can have utility in other contexts				
f. During today's session, students have had to reflect on which of their CREATE-skills they need to further develop to successfully complete the project				
g. During this session, students spent most of their time working on their projects				
h. I do not think that students' application of knowledge and skills within the context of this project reflects how these are applied in the real world				
i. My students feel ownership of this project: it is their project and their responsibility to complete it successfully				
j. While working on this project, students are developing their own best way of studying subject content for it to be of use towards project tasks and products				
k. It is clear to me how students' project activities help them develop specific CREATE-skills				
l. It is not clear to me how the content/knowledge that students study to complete their project links to the content/knowledge they need to learn to pass their exams				
m. It is clear to me how the subject matter studied by students aligns with what they need to know to complete this project successfully				

2. The following questions are also about **THE PROJECT SESSION THAT YOU HAVE JUST TAUGHT.**

How often do you think the following statements are accurate?

Question	Answer			
	Always	Most of the time	Most often not	Never
a. When I assisted students with their project work during this session, I offered solutions to the problems they encountered				
b. To complete their project tasks during today's session, students have had to use content from one or more subjects that are offered at this School (i.e. English, Science) to complete project tasks				
c. While working on this project during this session, students discovered independently where to find the information they need to complete the project				
d. During this session, students have provided/received feedback to/from one another on the products of their project activities				
e. During this session, students have received feedback feedback/critique on their project assignments from myself				
f. During this session, I instructed students about what they need to study to complete the project successfully				

3. The following questions are also about **THE PROJECT SESSION THAT YOU HAVE JUST TAUGHT.**

Do you agree with the following statements?

Question	Answer		
	Yes	No	Not sure
a. This project was presented to students with a clear explanation of the steps they needed to finish to complete their assignment			
b. During this session, students have worked or met with at least one adult from outside the school with relevant expertise/experience			

4. The following questions are about the work students have done in today's project session, and the things they have learned. Please answer these honestly and as completely as possible.

a. Please describe the subject content students have studied or learned **today** for this project and how you think it:

i) Aligns with the end product of the project

ii) Aligns with the curriculum content that is required for students to obtain their qualifications

Content and/or other studied or learned:	How it aligns with end product of the project:	How it aligns with content required for students' qualifications:

b. Please list the subject(s) students have had to use **today** while working on this project and what content or knowledge they have used

Subject (e.g. English)	Used knowledge (e.g., writing, running analyses, creating charts):

c. Please describe how the information, knowledge and/or skills students have obtained for this project **today** can, in your opinion, be useful in other contexts and situations

Information, Knowledge (e.g. English) or Skills (e.g. Communicating) obtained as part of project:	How it can be useful in other contexts:
---	---

5. If you would compare this project lesson to a lesson that is completely in line with 'protocol' or with a textbook version of how project-based learning is taught in your school:

a. Have you made any adaptations or other changes that make this lesson different from such a standard lesson? If yes, please list these. If no, please state this.

b. What was/were the reason(s) for any adaptations you have made?

Adaptations that make this lesson different from a standard PBL lesson: (e.g. skipped a part of the lesson, changed its content, etc)	Reason(s) for each adaptation made: (e.g. time constraints, misfit with the taught topic, etc)
--	---

Appendix B: Implementation checklists and questionnaires

B4: Checklist used by researcher for rating personal coaching implementation

1. Please rate the accuracy of these statements based on your observations of the current personal coaching session

How much do you agree with the following statements?

Question	Answer			
	Completely agree	Somewhat agree	Somewhat disagree	Completely disagree
a. During this coaching session, it was the coach who came up with solutions to the student's problems				
b. During this coaching session, the coach has advised student concretely on the actions she/he can undertake to achieve her/his goals				
c. During this coaching session, specified changes that are needed to improve the student's behaviours/learning were defined primarily by the student				
d. During this coaching session, the coach has provided sufficient opportunities for the student to discuss personal matters (such as experiences at home) without feeling exposed				
e. During this session, the coach has been sufficiently supportive of the student when she/he was in potentially uncomfortable situations				
f. During this coaching session, coach and student discussed student's attitudes towards himself and others, and coach used questions to help student self-examine these				

Do you agree with the following statements?

Question	Answer		
	Yes	No	Not sure
g. During this coaching session, student and coach have set academic and personal goals for student to work towards			
h. During this coaching session, student and coach have set actions that they each need to undertake for her/him to achieve her/his goals			
i. During this session, student and coach have reflected on the progress of actions they agreed towards achieving her/his goals			
j. During this coaching session, student and coach have reflected on students' academic progress			
k. During this coaching session, student and coach have discussed student's CREATE progress			
l. During this coaching session, student and coach have discussed what positively			

affects student's performance			
m. During this coaching session, student and coach have discussed what negatively affects student's performance			
n. During this coaching session, coach has guided student towards discovering methods for reinforcing her/his constructive thought patterns and emotions			
o. During this coaching session, coach has guided student towards discovering methods for reshaping dysfunctional thought patterns and emotions			
p. During this coaching session, student and coach discussed student's behaviour			
q. During this coaching session, student and coach looked at available information about student's performance in school (e.g., reports from coaches or test results) and used it to discuss the progress she/he has made			

2. Please rate the accuracy of these statements based on the qualitative data obtained from coach and student regarding the current session

How much do you agree with the following statements?

Question	Answer			
	Completely agree	Somewhat agree	Somewhat disagree	Completely disagree
a. Student is aware of her/his strengths with regards to her/his academic and CREATE progress				
b. Coach is aware of student's strengths with regards to her/his academic and CREATE progress				
c. Student is aware of areas for improvement with regards to her/his academic and CREATE progress				
d. Coach is aware of areas for improvement with regards to student's academic and CREATE progress				
e. Student is aware of her/his positive behaviour in school and at work				
f. Coach is aware of student's positive behaviour in school and at work				
g. Student is aware of the aspects of her/his behaviour in school and at work that she/he can improve				
h. Coach is aware of the aspects of student's behaviour in school and at work that she/he can improve				
i. Student is aware of her/his thought patterns that positively influence her/his perceptions of her/himself and others				
j. Coach is aware of student's thought patterns that positively influence her/his perceptions of her/himself and others				
k. Student is aware of the methods by which she/he student can				

reinforce her/his constructive thought patterns and emotions				
l. Student is aware of her/his thought patterns that negatively influence her/his perceptions of her/himself and others				
m. Coach is aware of student's thought patterns that negatively influence her/his perceptions of her/himself and others				
n. Student is aware of methods by which she/he can reshape dysfunctional thought patterns and emotions				
o. Students' academic and personal goals are clear to her/him				
p. Students' academic and personal goals are clear to the coach				
q. It is clear to student what actions coach and student have agreed to undertake in order for her/him to achieve her/his academic and personal goals				
r. It is clear to coach what actions student and coach have agreed to undertake in order for student to achieve her/his academic and personal goals				

Appendix B: Implementation checklists and questionnaires

B5: Questionnaire used by students for rating personal coaching implementation

INTRODUCTION

- ☆ **We are researchers and we would like to get your views about the personal coaching session you have just taken. This is NOT a test. There are no right or wrong answers!**
- ☆ **Your name is not on this questionnaire. This questionnaire is COMPLETELY CONFIDENTIAL. No-one will find out what you write other than the researchers. Your family, friends, coach and school will NOT find out what you write.**
- ☆ **The goal of this questionnaire is NOT to see how well you are doing in school or how well your coach does her or his job. We just want to see how your school is working with this important aspect of the Studio School model. No-one other than the researchers will find out what you write about yourself or your coach.**

COMPLETING THE QUESTIONNAIRE

- ☆ **Please answer all the questions as truthfully as you can. Please fill it all in and be honest. The answers you give will be used to learn more about how Studio Schools work.**
- ☆ **Please put a tick (✓) in the boxes provided. If you have made a mistake, put a cross (x) through the box and answer again.**

Thanks very much for your help!

Before you complete this questionnaire, please fill in your participant ID (given to you by the researcher) and today's date.

PARTICIPANT ID	
DATE	

1. These questions are about THE COACHING SESSION THAT YOU HAVE JUST FINISHED.

HOW MUCH DO YOU AGREE with the following statements?

Question	Answer			
	Totally Agree	Agree a bit	Don't really agree	Totally disagree
a. My academic and personal goals are clear to me				
b. I know ways to stay focused on thoughts that make me feel good and help me do well				
c. My coach does not know my strengths when it comes to my academic and CREATE progress				
d. My coach trusts me				
e. My coach is aware of aspects of my behaviour that I can improve				
f. In this coaching session, my coach was the one who came up with solutions to my problems				
g. My coach is aware of the aspects of my behaviour that are good				
h. My coach is aware of my thoughts that make me feel less good about myself and others				
i. I am aware of my strengths when it comes to my academic and CREATE progress				
j. I don't know how to reshape thoughts that make me feel bad and keep me from doing well				
k. In this coaching session, I don't think my coach gave me enough chances to talk about personal matters (like things that happen at home) without feeling exposed				
l. My coach is aware of things I can do better when it comes to my academic and CREATE progress				
m. I am not sure what my coach and I have agreed to do so that I achieve my academic and personal goals				
n. I don't know of any aspects of my behaviour that I can improve				

Question	Answer			
	Totally Agree	Agree a bit	Don't really agree	Totally disagree
o. I am aware of the thoughts that make me feel good about myself and others				
p. I know where I can do better when it comes to my academic and CREATE progress				
q. In this coaching session, my coach told me which things I should do to achieve my goals				
r. I am aware of the thoughts that make me feel less good about myself and others				
s. In this coaching session, my coach and I talked about how I see myself and others and she/he asked me questions to help me think deeper about it				
t. I trust my coach				
u. My coach doesn't understand me and my situation well enough				
v. My academic and personal goals are clear to my coach				
w. My coach is aware of my thoughts that make me feel good about myself and others				
x. In this coaching session, I was the one who came up with solutions to my problems				
y. I am aware of aspects of my behaviour that are good				
z. I don't feel that my coach has supported me enough in situations that may have made me feel uncomfortable during this session				

2. These questions are about THE COACHING SESSION THAT YOU HAVE JUST FINISHED.

DO YOU AGREE with the following statements?

Question	Answer		
	Yes	No	Don't know
a. In this coaching session, my coach and I have set actions that we need to do for me to achieve my goals			
b. In this coaching session, my coach and I talked about my behaviour			
c. In this coaching session, my coach has helped me to discover how to stay focused on thoughts that make me feel good and help me do well			
d. In this coaching session, my coach and I have talked about things that keep me from doing well			
e. In this coaching session, my coach and I have talked about my CREATE progress			
f. In this coaching session, my coach and I have set academic and personal goals for me to achieve			
g. In this coaching session, my coach and I have talked about things that help me to do well			
h. In this coaching session, my coach has helped me to discover how to reshape thoughts that make me feel bad and keep me from doing well			
i. In this coaching session, my coach and I have talked about my academic progress			
j. My coach and I have talked about the progress of the actions we agreed to do during earlier coaching sessions to achieve my goals			
k. During this coaching session, my coach and I looked at available information about how I do in school (for example: reports from my coaches or test results) and used it to discuss the progress I have made			

3. These questions are about the things you have agreed and done with your coach today. Please answer them honestly and as completely as possible.

a. Please describe your strengths and the areas you think you can improve for your academic and CREATE progress.

Strong academic aspects:	Strong CREATE skills:
Academic aspects needing improvement:	CREATE skills needing improvement:

b. Please describe the strengths and the areas you think you can improve with regards to your behaviour.

Strong aspects of behaviour:
Aspects of behaviour needing improvement:

c. Please describe:

i) *The thoughts that make you feel good about yourself and others, and the thoughts that make you feel less good about yourself and others*

ii) Please describe the ways by which you think you can boost thoughts that make you feel good and reshape thoughts that make you feel bad

Thoughts that make me feel good about myself and others:	Thoughts that make me feel less good about myself and others:
Ways I can boost these thoughts:	Ways I can reshape these thoughts:

d. Please list the academic and personal goals you have here.

Academic goals:

Personal goals:

e. Please describe the actions you and your have agreed to undertake to achieve your goals.

Actions you will do:	Actions your coach will do:
-----------------------------	------------------------------------

Appendix B: Implementation checklists and questionnaires

B6: Questionnaire used by coaches for rating personal coaching implementation

INTRODUCTION

- The University of Oxford and the Studio Schools Trust (SST) are working on a preliminary evaluation of the Studio School model. As part of this, we would like to ask you to complete a questionnaire regarding the personal coaching session you have just completed.
- This questionnaire is completely confidential. No party other than the researchers will see your answers. Your school and the Studio Schools Trust will not see your answers.
- It is NOT the purpose of this questionnaire to monitor your individual performance or to prescribe how to lead or structure a coaching session. Merely, the questionnaire seeks to gain insight into how your school works with this important aspect of the Studio School model. No other party than the researchers will see your individual answers or the answers your students give about your specific coaching sessions.

COMPLETING THE QUESTIONNAIRE

- Please complete all questions as truthfully as possible. Your answers will be used to learn more about the ways Studio Schools work with the Studio Schools model and the factors that may influence how they do so.
- Please put a tick (✓) in the boxes provided. In case you would like to rectify any mistakes, please put a cross (x) through the box and answer again.

Thank you very much for your help!

Before you complete this questionnaire, please fill in your participant ID (given to you by the researcher) and today's date.

PARTICIPANT ID	
DATE	

1. The following questions are about **the coaching session that you have just completed**.

How much do you agree with the following statements?

Question	Answer			
	Completely agree	Somewhat agree	Somewhat disagree	Completely disagree
a. It is clear to this student what her/his academic and personal goals are				
b. This student is aware of methods by which she/he can stay focused on her/his constructive thought patterns and emotions				
c. I am unaware of this student's strengths with regards to her/his academic and CREATE progress				
d. I trust this student				
e. I am aware of the aspects of this student's behaviour in school and at work that she/he can improve				
f. During this coaching session, I was the one who defined solutions to my student's problems				
g. I am aware of the positive aspects of this student's behaviour				
h. I am aware of this student's thought patterns that negatively influence her/his perceptions of her/himself and others				
i. This student is aware of her/his strengths with regards to her/his academic and CREATE progress				
j. This student is unaware of methods by which she/he can reshape dysfunctional thought patterns and emotions				
k. During this coaching session, I do not think I have provided sufficient opportunities for my student to discuss personal matters (such as experiences at home) without feeling exposed				
l. I am aware of this student's areas for improvement with regards to their academic and CREATE progress				
m. It is not clear to me what actions my student and I have agreed to undertake in order for her/him to achieve her/his academic and personal goals				
n. This student is unaware of the aspects of her/his behaviour in school and at work that she/he can improve				

Question	Answer			
	Completely agree	Somewhat agree	Somewhat disagree	Completely disagree
o. This student is aware of this her/his thought patterns that positively influence her/his perceptions of her/himself and others				
p. This student is aware of her/his areas for improvement with regards to their academic and CREATE progress				
q. During this coaching session, I instructed my student which actions she/he should undertake to achieve her/his goals				
r. This student is aware of her/his thought patterns that negatively influence her/his perceptions of her/himself and others				
s. During this coaching session, my student and I talked about her/his attitudes towards her/himself and others, and I used questions to help my student self-examine these				
t. This student trusts me				
u. I do not understand this student and their situation well enough				
v. The academic and personal goals of this student are clear to me				
w. I am aware of this student's thought patterns that positively influence her/his perceptions of her/himself and others				
x. During this coaching session, my student was the one who defined solutions to her/his problems				
y. This student is aware of the positive aspects of her/his behaviour				
z. I do not feel that I have been supportive enough of this student when she/he was in potentially uncomfortable situations during this session				

2. The following questions are also about **the coaching session that you have just completed**.

Do you agree with the following statements?

Question	Answer		
	Yes	No	Not sure
a. During this coaching session, my student and I have set actions that she/he and I need to undertake for her/him to achieve her/his goals			
b. During this coaching session, my student and I discussed her/his behaviour			
c. During this coaching session, I have guided my student towards discovering methods for focusing on her/his constructive thought patterns and emotions			
d. During this coaching session, my student and I have discussed what negatively affects her/his performance			
e. During this coaching session, my student and I have discussed her/his CREATE progress			
f. During this coaching session, my student and I have set academic and personal goals for her/him to work towards			
g. During this coaching session, my student and I have discussed what positively affects her/his performance			
h. During this coaching session, I have guided my student towards discovering methods for reshaping dysfunctional thought patterns and emotions			
i. During this coaching session, my student and I have reflected on her/his academic progress			
j. My student and I have reflected on the progress of actions we agreed during earlier coaching sessions towards achieving her/his goals			
k. During this coaching session, my student and I looked at available information about their performance in school (e.g., reports from coaches or test results) and used it to discuss the progress she/he has made			

3. The following questions are about the student you have just coached, and the things you and your student have done and agreed to during this coaching session. Please answer these honestly and as completely as possible.

a. Please describe this student's strengths and weaknesses with regards to their academic and CREATE progress

Strong academic aspects:	Strong CREATE skills:
Academic aspects needing improvement:	CREATE skills needing improvement:

b. Please describe this student's weaknesses with regards to their behaviour in school and at work

Strong aspects of behaviour:
Aspects of behaviour needing improvement:

c. Please describe:

i) This student's thought patterns that influence her/his perceptions of her/himself and others positively and negatively

ii) The specific methods by which this student can reinforce her/his constructive thought patterns and emotions and reshape dysfunctional ones.

Positive thought patterns of this student: Methods through which these can be reinforced:	Negative thought patterns of this student: Methods through which these can be reshaped:
--	--

d. Please describe, according to your view, the current academic and personal goals of this student

Academic goals:

Personal goals:

e. Please describe the actions you and your student have agreed to undertake to achieve her/his goals

Actions undertaken by yourself:	Actions undertaken by your student:

4. If you would compare this coaching session to a session that is completely in line with 'protocol' or a textbook version of how personal coaching is being done in your school:

a. Have you made any adaptations or other changes that make this session different from such a standard session? If yes, please list these. If no, please state this.

b. What was/were the reason(s) for any adaptations you have made?

Adaptations that make this session different from a standard session: <i>(e.g. skipped certain discussion topics or actions, changed the approach, etc.)</i>	Reason(s) for each adaptation made: <i>(e.g. time constraints, misfit with this student's needs, etc)</i>

Appendix C: Results of psychometric properties assessments

C1. Polychoric correlations for PBL items: first EFA

	AA2	AA5	FA2	FA3	FA4	FA5	AB1	AB2	FB1	AB3	FB2	AC1	FC1	AD1	FD1	FD2	AE1A	AE1B	AE3	FE1	FE2	FE3	FE4	FF1
AA2	1.00																							
AA5	-0.03	1.00																						
FA2	-0.52	-0.08	1.00																					
FA3	-0.33	-0.15	0.46	1.00																				
FA4	-0.15	0.12	0.15	0.06	1.00																			
FA5	-0.38	-0.01	0.45	0.31	0.08	1.00																		
AB1	0.47	-0.07	-0.51	-0.28	-0.15	-0.40	1.00																	
AB2	0.39	-0.24	-0.35	-0.13	-0.12	-0.20	0.49	1.00																
FB1	-0.26	-0.12	0.41	0.42	0.16	0.22	-0.31	-0.17	1.00															
FB2	-0.19	0.11	0.43	0.21	0.12	0.22	-0.50	-0.37	0.32	1.00														
FB3	-0.29	0.09	0.45	0.35	0.27	0.41	-0.44	-0.22	0.46	0.36	1.00													
AC1	-0.34	-0.13	0.38	0.42	0.11	0.35	-0.44	-0.48	0.36	0.37	0.28	1.00												
FC1	-0.21	0.11	0.08	0.16	0.47	0.11	-0.08	-0.20	0.08	0.19	0.20	0.10	1.00											
AD1	-0.10	0.00	0.29	0.21	0.09	0.27	-0.21	-0.22	0.17	0.19	0.25	0.05	0.10	1.00										
FD1	-0.18	-0.13	0.37	0.21	-0.11	0.32	-0.35	-0.34	0.29	0.27	0.34	0.32	-0.08	0.48	1.00									
FD2	-0.46	-0.03	0.53	0.43	0.13	0.48	-0.44	-0.34	0.37	0.28	0.49	0.32	0.22	0.46	0.57	1.00								
AE1A	-0.40	0.07	0.35	0.24	0.14	0.20	-0.39	-0.42	0.24	0.41	0.29	0.33	0.06	0.21	0.33	0.30	1.00							
AE1B	-0.33	-0.01	0.46	0.27	0.19	0.27	-0.60	-0.50	0.30	0.39	0.32	0.45	0.06	0.20	0.23	0.30	0.47	1.00						
AE3	-0.32	0.30	0.28	0.14	0.29	0.32	-0.40	-0.40	0.24	0.45	0.37	0.38	0.14	0.17	0.33	0.40	0.36	0.31	1.00					
FE1	-0.16	0.12	0.03	0.09	0.27	0.11	-0.11	-0.12	0.05	0.09	0.14	0.01	0.22	0.15	0.12	0.12	0.06	0.04	0.27	1.00				
FE2	-0.35	0.09	0.32	0.50	0.01	0.37	-0.39	-0.14	0.21	0.29	0.47	0.21	0.17	0.21	0.26	0.45	0.21	0.44	0.21	0.18	1.00			
FE3	-0.48	-0.01	0.45	0.34	0.07	0.38	-0.35	-0.44	0.37	0.31	0.53	0.29	0.16	0.31	0.36	0.57	0.31	0.22	0.21	0.09	0.52	1.00		
FE4	-0.33	-0.03	0.49	0.36	0.03	0.45	-0.35	-0.45	0.18	0.34	0.45	0.49	0.18	0.15	0.39	0.42	0.34	0.33	0.20	0.17	0.41	0.44	1.00	
FF1	-0.34	-0.12	0.32	0.29	0.35	0.22	-0.20	-0.16	0.18	0.16	0.45	0.11	0.33	0.07	0.03	0.27	0.06	0.24	0.02	0.26	0.31	0.35	0.35	1.00

Appendix C: Results of psychometric properties assessments

C2: Polychoric correlations for PBL items: repeated EFA

	AA2	AB1	AB2	AC1	AE1A	AE1B	AE3	AD1	FA2	FA3	FA5	FB1	FB2	FB3	FD1	FD2	FE2	FE3	FE4	FA4	FC1	FF1
AA2	1.00																					
AB1	0.47	1.00																				
AB2	0.39	0.49	1.00																			
AC1	0.34	0.44	0.48	1.00																		
AE1A	0.40	0.39	0.42	0.33	1.00																	
AE1B	0.33	0.60	0.50	0.45	0.47	1.00																
AE3	0.32	0.40	0.40	0.38	0.36	0.31	1.00															
AD1	0.10	0.21	0.22	0.05	0.21	0.20	0.17	1.00														
FA2	0.52	0.51	0.35	0.38	0.35	0.46	0.28	0.29	1.00													
FA3	0.33	0.28	0.13	0.42	0.24	0.27	0.14	0.21	0.46	1.00												
FA5	0.38	0.40	0.20	0.35	0.20	0.27	0.32	0.27	0.45	0.31	1.00											
FB1	0.26	0.31	0.17	0.36	0.24	0.30	0.24	0.17	0.41	0.42	0.22	1.00										
FB2	0.19	0.50	0.37	0.37	0.41	0.39	0.45	0.19	0.43	0.21	0.22	0.32	1.00									
FB3	0.29	0.44	0.22	0.28	0.29	0.32	0.37	0.25	0.45	0.35	0.41	0.46	0.36	1.00								
FD1	0.18	0.35	0.34	0.32	0.33	0.23	0.33	0.48	0.37	0.21	0.32	0.29	0.27	0.34	1.00							
FD2	0.46	0.44	0.34	0.32	0.30	0.30	0.40	0.46	0.53	0.43	0.48	0.37	0.28	0.49	0.57	1.00						
FE2	0.35	0.39	0.14	0.21	0.21	0.44	0.21	0.21	0.32	0.50	0.37	0.21	0.29	0.47	0.26	0.45	1.00					
FE3	0.48	0.35	0.44	0.29	0.31	0.22	0.21	0.31	0.45	0.34	0.38	0.37	0.31	0.53	0.36	0.57	0.52	1.00				
FE4	0.33	0.35	0.45	0.49	0.34	0.33	0.20	0.15	0.49	0.36	0.45	0.18	0.34	0.45	0.39	0.42	0.41	0.44	1.00			
FA4	0.15	0.15	0.12	0.11	0.14	0.19	0.29	0.09	0.15	0.06	0.08	0.16	0.12	0.27	-0.11	0.13	0.01	0.07	0.03	1.00		
FC1	0.21	0.08	0.20	0.10	0.06	0.06	0.14	0.10	0.08	0.16	0.11	0.08	0.19	0.20	-0.08	0.22	0.17	0.16	0.18	0.47	1.00	
FF1	0.34	0.20	0.16	0.11	0.06	0.24	0.02	0.07	0.32	0.29	0.22	0.18	0.16	0.45	0.03	0.27	0.31	0.35	0.35	0.35	0.33	1.00

Appendix C: Results of psychometric properties assessments

C3: Polychoric correlations for personal coaching items: EFA

	AA1A	AA1B	AA2	AA3A	AA3B	AA4A	AA4B	FA1A	FA1B	FA1C	FA1D	FA2A	FA2B	FA2C	FA2D	FA3A	FA3B	FA3C	FA3D	FA4A	FA4B	AB1	AB2	AB3
AA1A	1.00																							
AA1B	0.45	1.00																						
AA2	0.31	0.02	1.00																					
AA3A	-0.01	0.02	0.57	1.00																				
AA3B	0.29	0.38	0.14	0.21	1.00																			
AA4A	0.22	0.30	0.11	0.58	0.64	1.00																		
AA4B	-0.03	-0.04	0.27	0.57	0.37	0.47	1.00																	
FA1A	0.44	0.25	0.18	0.29	0.33	0.52	0.19	1.00																
FA1B	0.41	0.53	-0.02	-0.11	0.40	0.23	-0.13	0.48	1.00															
FA1C	0.44	0.59	-0.15	-0.18	0.20	0.02	-0.36	0.14	0.58	1.00														
FA1D	0.35	0.45	-0.12	0.05	0.36	0.37	0.16	0.29	0.53	0.65	1.00													
FA2A	0.30	0.42	0.03	0.34	0.42	0.59	0.03	0.32	0.39	0.57	0.53	1.00												
FA2B	0.38	0.49	-0.14	-0.25	0.40	0.13	-0.22	0.36	0.49	0.59	0.43	0.42	1.00											
FA2C	-0.12	0.21	-0.14	0.07	0.20	0.19	0.16	0.28	0.29	0.22	0.25	0.42	0.35	1.00										
FA2D	0.08	0.32	-0.14	-0.04	0.47	0.23	-0.12	0.18	0.57	0.49	0.43	0.53	0.48	0.35	1.00									
FA3A	0.38	0.24	0.14	0.29	0.22	0.51	0.12	0.42	0.21	0.13	0.22	0.55	0.35	0.23	0.41	1.00								
FA3B	0.15	0.38	-0.08	-0.17	-0.21	-0.15	-0.36	0.18	0.51	0.65	0.59	0.23	0.38	0.15	0.40	0.05	1.00							
FA3C	0.33	0.34	0.22	0.16	0.34	0.28	0.18	0.22	0.37	0.37	0.54	0.36	0.33	0.13	0.37	0.32	0.35	1.00						
FA3D	0.16	0.25	0.00	0.14	0.31	0.42	0.02	0.38	0.50	0.33	0.49	0.45	0.19	0.21	0.63	0.43	0.52	0.44	1.00					
FA4A	0.07	0.08	-0.22	-0.08	0.24	0.30	-0.18	0.24	0.33	0.52	0.46	0.41	0.36	0.27	0.39	0.17	0.43	0.27	0.41	1.00				
FA4B	0.17	0.24	-0.02	0.10	0.34	0.29	-0.05	0.36	0.35	0.42	0.46	0.29	0.46	0.25	0.29	0.14	0.27	0.25	0.30	0.45	1.00			
AB1	0.06	0.46	-0.02	0.40	0.50	0.70	0.28	0.37	0.45	0.38	0.53	0.65	0.42	0.26	0.44	0.35	0.25	0.27	0.41	0.37	0.27	1.00		
AB2	0.06	0.46	-0.02	0.40	0.50	0.70	0.28	0.37	0.45	0.38	0.53	0.65	0.42	0.26	0.44	0.35	0.25	0.27	0.41	0.37	0.27	0.91	1.00	
AB3	-0.09	0.30	0.00	0.26	0.54	0.42	0.30	0.25	0.40	0.19	0.44	0.45	0.57	0.48	0.51	0.26	0.22	0.23	0.34	0.27	0.32	0.65	0.65	1.00
FB1A	0.01	-0.08	-0.26	-0.12	-0.09	0.15	-0.24	0.47	0.31	0.26	0.38	0.37	0.43	0.37	0.31	0.25	0.55	0.01	0.44	0.56	0.33	0.35	0.35	0.36
FB1B	0.46	0.59	0.05	0.10	0.50	0.41	-0.04	0.37	0.56	0.62	0.53	0.72	0.56	0.32	0.51	0.39	0.34	0.40	0.37	0.38	0.40	0.56	0.56	0.51
FB2STU	0.09	0.50	0.08	0.13	0.55	0.34	-0.04	0.33	0.53	0.48	0.40	0.46	0.60	0.35	0.53	0.13	0.35	0.48	0.29	0.38	0.35	0.58	0.58	0.57
AC1	0.50	0.47	0.03	0.04	0.41	0.34	0.15	0.30	0.39	0.32	0.42	0.46	0.39	0.19	0.47	0.38	0.20	0.45	0.39	0.27	0.24	0.28	0.28	0.33
AC2	-0.09	-0.44	-0.03	-0.12	-0.26	-0.30	-0.13	-0.35	-0.27	-0.17	-0.26	-0.25	-0.47	-0.15	-0.16	-0.12	-0.28	-0.17	-0.08	-0.28	-0.14	-0.56	-0.56	-0.47
FC1A	-0.19	-0.39	-0.07	-0.11	-0.07	-0.14	-0.16	0.02	-0.07	-0.13	-0.07	0.01	-0.21	0.10	0.00	-0.12	-0.08	-0.12	0.07	0.09	0.12	-0.26	-0.26	-0.19
FC1B	0.05	0.14	0.09	0.11	0.15	0.13	0.21	0.08	0.14	0.04	0.26	0.20	0.01	0.12	0.28	0.24	0.17	0.29	0.20	-0.03	-0.06	0.17	0.17	0.14
AD1	0.15	0.51	-0.18	-0.16	0.19	0.19	-0.17	0.29	0.47	0.53	0.21	0.54	0.41	0.34	0.49	0.36	0.34	0.03	0.36	0.29	0.18	0.43	0.43	0.24
FD1A	0.48	0.42	-0.22	-0.17	0.37	0.21	-0.13	0.54	0.63	0.72	0.54	0.63	0.74	0.41	0.53	0.28	0.40	0.23	0.26	0.48	0.23	0.51	0.51	0.41
FD1B	0.42	0.31	-0.01	-0.05	0.43	0.26	-0.16	0.51	0.65	0.48	0.24	0.58	0.72	0.39	0.70	0.56	0.25	0.22	0.46	0.33	0.30	0.43	0.43	0.46
FD2	0.43	0.56	0.21	0.28	0.52	0.44	0.25	0.39	0.44	0.52	0.43	0.58	0.60	0.35	0.48	0.47	0.29	0.34	0.38	0.34	0.34	0.54	0.54	0.49
FD3	0.56	0.58	0.09	0.05	0.38	0.22	-0.01	0.32	0.50	0.74	0.54	0.61	0.65	0.29	0.53	0.31	0.47	0.28	0.35	0.32	0.36	0.41	0.41	0.42
LOOSE1Q	-0.05	0.53	-0.30	0.08	-0.02	0.20	-0.12	-0.02	0.01	0.31	0.19	0.34	0.36	0.07	0.05	0.24	0.19	0.17	-0.08	0.15	0.21	0.36	0.36	0.24

Appendix C: Results of psychometric properties assessments

C3: Polychoric correlations for personal coaching items: EFA, continued

	FB1A	FB1B	FB2ST	AC1	AC2	FC1A	FC1B	AD1	FD1A	FD1B	FD2	FD3	LOOSE.IQ
AA1A													
AA1B													
AA2													
AA3A													
AA3B													
AA4A													
AA4B													
FA1A													
FA1B													
FA1C													
FA1D													
FA2A													
FA2B													
FA2C													
FA2D													
FA3A													
FA3B													
FA3C													
FA3D													
FA4A													
FA4B													
AB1													
AB2													
AB3													
FB1A	1.00												
FB1B	0.34	1.00											
FB2STU	0.19	0.59	1.00										
AC1	0.15	0.54	0.34	1.00									
AC2	-0.34	-0.35	-0.47	-0.14	1.00								
FC1A	0.29	-0.07	-0.17	0.00	0.35	1.00							
FC1B	0.04	0.22	0.14	0.25	-0.03	0.14	1.00						
AD1	0.27	0.59	0.43	0.26	-0.21	-0.16	0.15	1.00					
FD1A	0.50	0.60	0.48	0.46	-0.40	-0.04	0.09	0.47	1.00				
FD1B	0.45	0.55	0.51	0.53	-0.23	-0.02	0.04	0.56	0.77	1.00			
FD2	0.19	0.55	0.49	0.50	-0.38	-0.14	0.17	0.47	0.64	0.63	1.00		
FD3	0.23	0.62	0.52	0.54	-0.27	-0.23	0.14	0.61	0.73	0.64	0.74	1.00	
LOOSE.IQ	0.02	0.34	0.36	0.31	-0.27	-0.20	0.09	0.30	0.16	0.08	0.30	0.21	1.00

Appendix C: Results of psychometric assessments

C4: Overview of PBL structural equation models compared in CFA

Model 1 (Alternative discussed in Chapter 5)

Four factor model that is based exactly on the EFA output, with one function (FF1) integrated into an activities factor, one activity (AA2) integrated into a function factor, and the CREATE skills items (AD1, FD1, FD2) forming a separate factor. Model passed all fit tests and yielded best overall estimates (small improvement when compared with PBL.model3, yet theoretically inconsistent)

Chi-squared p : 0.113

CFI: 0.995

TLI: 0.994

RMSEA: 0.000 – 0.052, point estimate 0.032, p RMSA < 0.05 = 0.933

Model 2 (Chosen model)

The chosen model with 1 factor exclusively comprising activities and 3 factors exclusively comprising functions, with one separate activity item and function FF1 removed. Model passed all fit tests, and yielded best overall estimates when these are seen as a continuum.

Chi-Squared p : 0.352

CFI: 0.999

TLI: 0.998

RMSEA: 0.000 – 0.045, point estimate 0.018, p RMSA < 0.05 = 0.979

Model 3 (Discarded)

Three factor model with one activity factor and two function factors. FD1 and FD2 are integrated into a main function factor. Model failed Chi-square and RMSEA tests of fit but came close to passing.

Chi-squared p : 0.032

CFI: 0.992

TLI: 0.991

RMSEA: 0.013 – 0.061, point estimate 0.042, p RMSA < 0.05 = 0.748

Model 4 (Discarded)

The original model with 1 activity factor and 3 function factors, from which AD1 is left out as a separate item. FF1 is included. Model failed Chi-squared fit and RMSEA fit test:

Chi-Squared p : 0.004

CFI: 0.988

TLI: 0.986

RMSEA: 0.029 – 0.066, p RMSEA < 0.05 = 0.516

Model 5 (Discarded)

Two factor model where all activities and functions except for FA4 and FC1 are grouped together in one factor. Model failed Chi-squared fit and RMSEA fit test:

Chi-squared p : 0.000

CFI: 0.981

TLI: 0.979

RMSEA: 0.045 – 0.078, point estimate 0.062, p RMSA < 0.05 = 0.118

Model 6 (Discarded)

Three factor model with FB2 integrated into a mostly activity Factor, AD1 integrated into a mostly function factor, and FF1 integrated into a separate factor. Model failed Chi-squared and RMSEA fit tests.

Chi-squared p : 0.000

CFI: 0.984

TLI: 0.982

RMSEA: 0.038-0.071, point estimate 0.055, p RMSA $<0.05=0.282$

Appendix D: Correlation scatter plots

D1. Scatter plots depicting session-level correlations for PBL fidelity scores

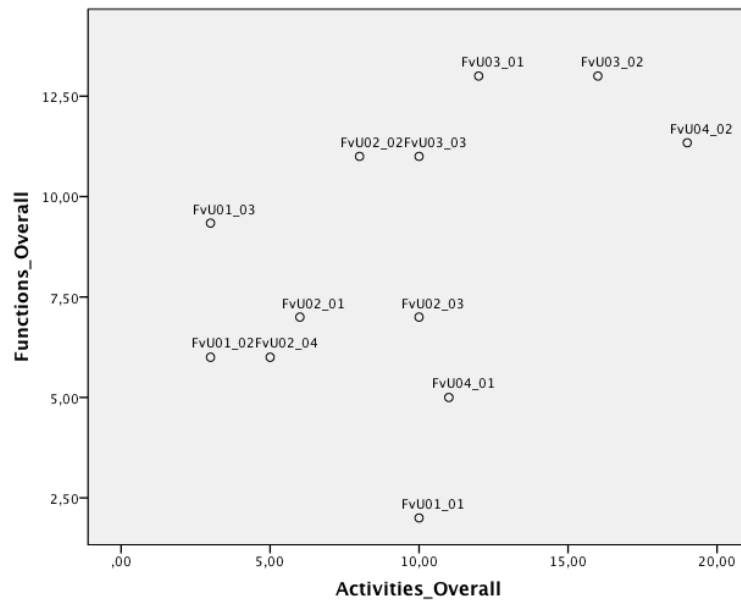


Figure D1.1: Scatter plot depicting session-level correlations for PBL fidelity scores as rated by the researcher

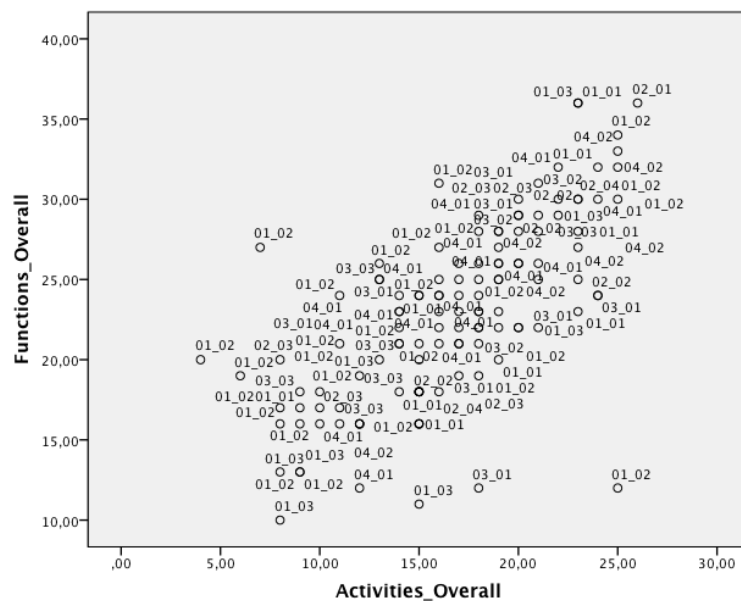


Figure D1.2: Scatter plot depicting session-level correlations for PBL fidelity scores as rated by the students

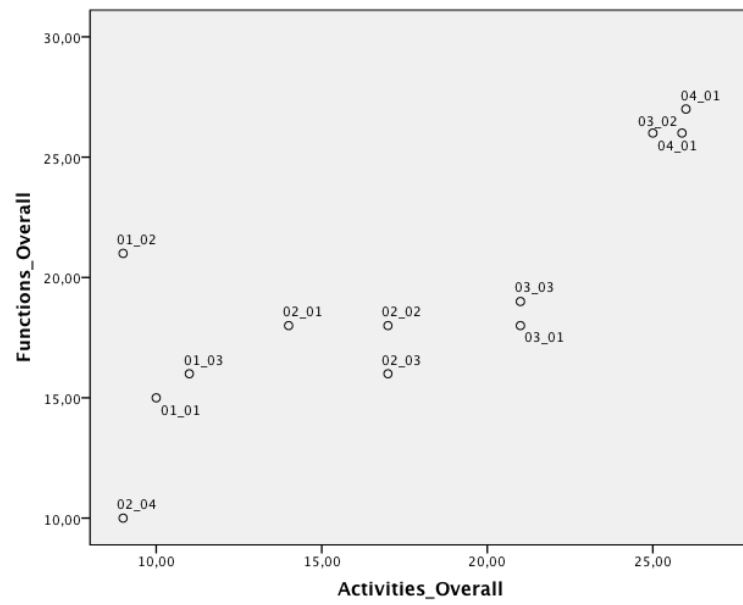


Figure D1.3: Scatter plot depicting session-level correlations for PBL fidelity scores as rated by the teachers

Appendix D: Correlation scatter plots

D2. Scatter plots depicting session-level correlations for personal coaching fidelity scores

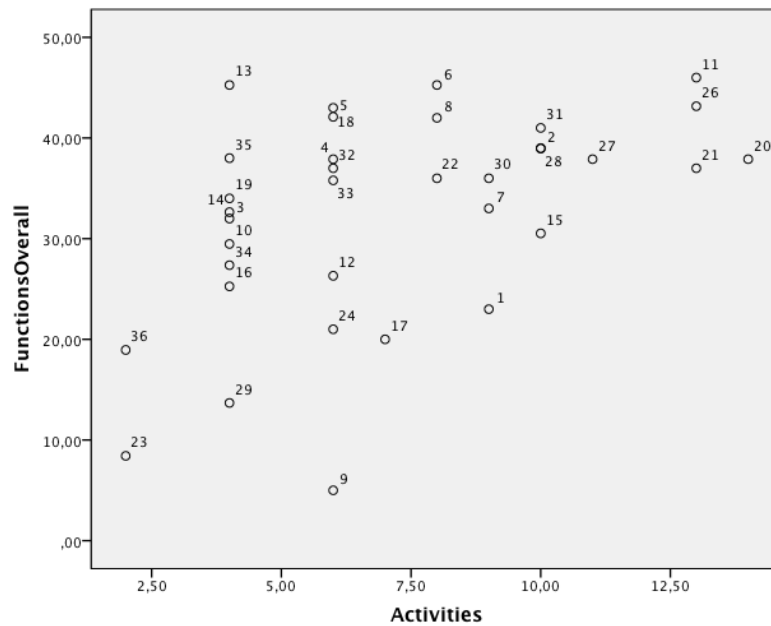


Figure D2.1: Scatter plot depicting session-level correlations for personal coaching fidelity scores as rated by the researcher

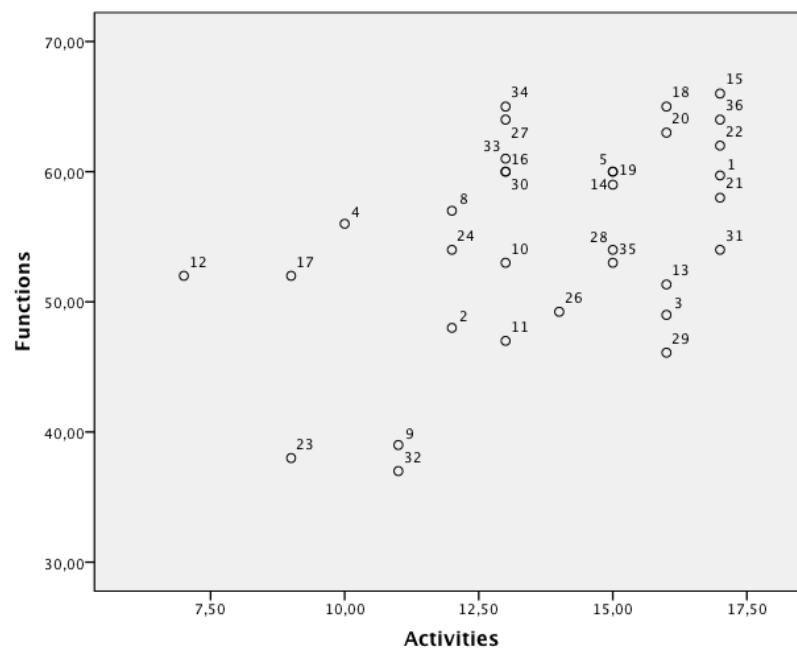


Figure D2.2: Scatter plot depicting session-level correlations for personal coaching fidelity scores as rated by the students

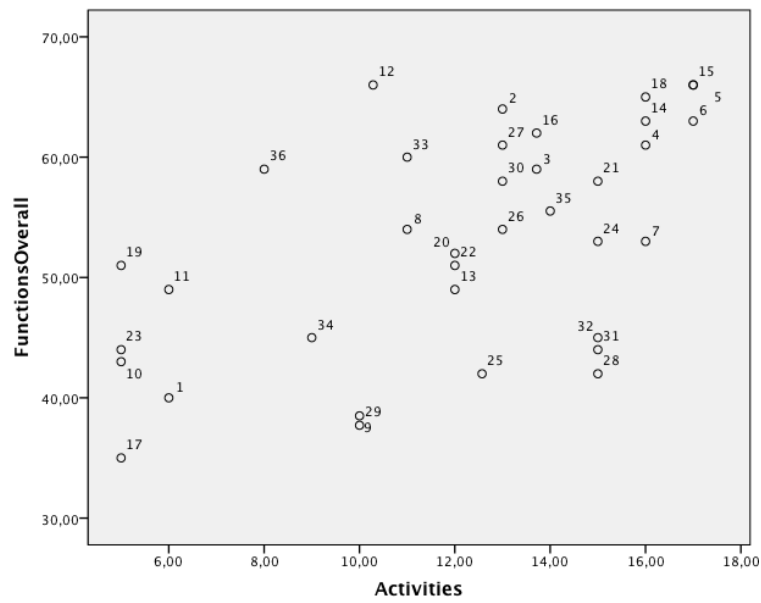


Figure D2.3: Scatter plot depicting session-level correlations for personal coaching fidelity scores as rated by the coaches