

**Fostering and Capturing Children's Inner
Motivation to Learn in the Early Primary
Classroom in England**

Soizic Le Courtois

Darwin College

University of Cambridge

This thesis is submitted for the degree of Doctor of Philosophy.

January 2022

Declaration

This thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text.

I further state that no substantial part of my thesis has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text.

It does not exceed the prescribed word limit for the relevant Degree Committee.

This thesis was funded by the LEGO Foundation, Cambridge Trust, and funds of the Stepping Stones project of the Centre for Research on Play in Education, Development and Learning (PEDAL).

Abstract

Fostering and Capturing Children's Inner Motivation to Learn in the Early Primary Classroom in England

Soizic Le Courtois

Children are born naturally curious and eager to learn, but as they go through school this inner motivation to learn diminishes. Yet children's inner motivation to learn is essential to deep learning, positive attitudes to school and wellbeing. Self-Determination Theory suggests that supporting children's need for autonomy – that is to say the feeling that actions stem from internal sources rather than being imposed externally – is essential to supporting inner motivational resources. This thesis is concerned with how teachers may be able to support children's autonomy and inner motivation in the early Primary classroom in England and how we may be able to capture changes in children's inner motivation in those settings. It is divided into two parts.

In Part I, I used interpretive methods to understand teachers' attempts to provide greater opportunities for children's autonomy in Year 1 classrooms through a professional development programme. This programme was developed by a team of researchers at the PEDAL centre using a Community of Practice model and involved nine teachers in trying out strategies to support children's autonomy. Through stories of change, I show that teachers' use and interpretations of the strategies varied, and this was affected by the teachers' school context and their own beliefs. Through thematic analysis, I show that the classrooms in the study functioned as ecosystems of teacher control, which was itself under pressure from top-down directions through governmental policies and institutions as well as senior leaders. This resulted in a teaching mindset focused on strict learning objectives which left little space for children to take ownership of their learning. Despite this, teachers were sometimes able to provide pockets of space for children's autonomy, though these took diverse forms. The extent of these spaces for autonomy depended on individual school and classroom contexts. The proposed model – pockets of space within an ecosystem of teacher control – explains the tensions between teachers' need for control in the classroom and opportunities for children's autonomy, as well as areas where teachers' attempts to increase children's autonomy were successful. In particular, I show that teachers needed to provide support and stimulation as well as space in order to support both autonomy and inner motivation.

Part II is concerned with measuring inner motivation for research purposes and in particular for future evaluations of the above professional development programme. This research focuses on the validity and reliability of an existing instrument, the Leuven Involvement Scale (LIS). This instrument aims to capture a form of engagement in learning activities that is related to inner motivation. The studies in Part II investigate the reliability and stability of the instrument, as well as factors associated with variation in engagement using multilevel modelling. I found that the LIS can be reliable as long as raters share a common understanding of different child behaviours in the classroom. In addition, I found that engagement varied hugely from one moment to the next, with very little variation between children. What little variation existed between children was explained by the association between engagement and aspects of children's self-regulatory capacities, namely effortful control and negative emotions, measured through the Strengths and Difficulties Questionnaire (SDQ) and Child Behaviour Questionnaire (CBQ). However, overall this research suggests that it is the individual moment that matters, rather than characteristics of the children. To better understand the influence of contextual factors, I investigated the association of activity setting (whether children are in teacher-directed, independent or free choice situations) with engagement. Children were significantly more engaged in free choice settings compared to whole class teacher-directed settings. However, there was a large amount of remaining variation and I discuss the implications this has for the role of teachers in supporting children's engagement.

Overall, this thesis makes a contribution towards our understanding of children's autonomy and inner motivation in the classroom and teaching practices that support it, as well as how we may be able to study it in classroom contexts.

Preface

In certifying that this thesis is my own work, I list here the publications in which some of this work has appeared. I certify that I was the author of all writing which appears in this thesis.

Baker, S. T., Le Courtois, S., & Eberhart, J. (2021). Making space for children's agency with playful learning. *International Journal of Early Years Education*.

<https://doi.org/10.1080/09669760.2021.1997726>

Blaskova, L. J., Le Courtois, S., Baker, S. T., Gibson, J. L., Ramchandani, P. G., O'Farrelly, C., & Fink, E. (2020). Participant engagement with play research – examples and lessons learned from the Centre for Play in Education, Development and Learning. *International Journal of Play*, 9(4), 365–381. <https://doi.org/10.1080/21594937.2020.1843804>

Le Courtois, S. & Baker, S. (2020, Apr 17 - 21) *Stepping-Stones for Children's Autonomy: A Community of Practice Project in English Classrooms* [Structured Poster Session]. AERA Annual Meeting San Francisco, CA <http://tinyurl.com/y3zv9xg8> (Conference Cancelled)

Acknowledgements

First and foremost, I would like to thank my supervisor, Prof. Sara Baker, for her advice, support, care and pep-up talks over the past five years. Writing a thesis drawing on Self-Determination Theory, I was perhaps particularly aware of the need for autonomy, competence and relatedness, and I am so lucky that Sara supported these needs so generously. To feel trusted and valued in this way is an experience that I will carry with me for the rest of my career. I am also grateful for her understanding of my circumstances – becoming a student parent at the start of my PhD – and the practical difficulties of juggling parenting and research, in particular in times of pandemic.

I am grateful to the LEGO Foundation and Cambridge Trust for funding this research and believing I was the right candidate to see it through. I am also thankful for the PEDAL research centre, both for supporting my research and for the welcoming, playful community of researchers I was fortunate enough to be part of. In particular, I am thankful to Anna Vidos for her friendship and administrative helping hand and to fellow PhD students Krishna Kulkarni and Emily Goodacre for never giving up on asking me to come along to social events, despite my terrible track record. I also wish to thank Prof. Paul Ramchandani for always being on hand to answer questions, for his unwavering optimism and for leading such a wonderful research centre to work in – such things do not happen by accident. I am also grateful to my advisor Dr Jenny Gibson for her continuous support and advice, and in particular for stepping in as interim supervisor during my own supervisor's maternity leave. Special thanks also to Dr Pablo Torres for helping with initial reliability coding in schools and for introducing me to Krippendorff's Alpha (and an easy way to calculate it), undoubtedly saving me hours of work. A huge thank you must go to the Stepping Stones team, and in particular to Hayley Gains for her extensive help with data collection and to Dr Audrey Kittredge, for her friendship and support, for the many conversations we had about this thesis, and for her wonderful conference networking skills. These thanks also extend to the more recently formed TRAIL team, and I am particularly indebted to Dee Rutgers for joining my Leuven Scale bandwagon and helping with reliability coding, as well as to Allison Haack for the wonderful graphics she provided for this thesis.

Thank you also to the broader education community in Cambridge and beyond, and in particular to Dr Elaine Wilson as well as fellow students Nomisha Kurian, Tania Clarke, Gabby Arengé and Jwalin Patel for the conversations on methodology, motivation, wellbeing, agency and education in general. Thank you for making time to listen and chat – it

meant a lot. Thank you also to Prof. Nancy Perry for taking the time to tell me about her research and use of running records. I am also incredibly thankful for Dr Joanne Waterhouse's advice over six years ago to consider academia, and for believing that the questions I had about education were worth pursuing.

I also thank all the teachers and children who entrusted me with their experiences and thoughts. I am incredibly grateful to have been invited into these classrooms and it has been a privilege to be able to sit in and observe so many lessons. This research would have been impossible without their participation, and I hope I have made it worthwhile for them.

I am also thankful for the wonderful support network I have around me and without which this PhD would have been much more difficult and less pleasant – to my in-laws for the countless hours of childcare, and to friends and neighbours for providing such a lovely community to live in: to Jess and Kirsty, for the regular lockdown toddler walks and chats, and the VIE crowd: Nigel, Liz, Xieanne, San, Hellen, and all the kids – thank you for helping me keep my sanity during what has been quite tough times.

I am also forever grateful to my parents, Patrick and Isabelle Le Courtois, for their support and pride in all I do. I also owe them my most deep-seated beliefs about childhood and education – and in particular the importance of trusting and respecting children. This thesis and I would not be the same if my upbringing had been different. Thank you also to my sisters, who are always only a video call away in time of need. I also wish to thank my young son, for understanding that I sometimes had to shut myself away to work and could not play (and for putting up with it when he did not understand). And most importantly, I am incredibly grateful to my husband, Mark Lewis, for his love and unwavering faith in me and this project, despite the impact it has had on our lives as a family. His steadfastness in this has meant more to me than words can express.

Table of Contents

Declaration.....	iii
Abstract.....	iv
Preface.....	vi
Acknowledgements.....	vii
Table of Contents.....	ix
List of Tables.....	xiv
List of Figures.....	xv
List of Abbreviations.....	xvi
Preamble - Some Personal Introductory Remarks.....	1
Chapter 1 - Introduction.....	3
1.1 Thesis Aims.....	3
1.2 Theoretical Underpinning: Self-Determination Theory and Autonomous Motivation ...	4
1.3 Rationale for the Thesis.....	5
1.4 Situating the Thesis.....	8
1.5 Summary of Research Aims.....	14
Chapter 2 - Review of the Literature on Fostering Children’s Inner Motivation by Supporting Their Autonomy in the Classroom.....	15
2.1 Autonomy as Supportive of Inner Motivation.....	15
2.2 Autonomy and Choice in Child-Centred Pedagogies.....	21
2.3 Autonomy as a Need and a Right.....	25

2.4 Conclusion to Chapter 2.....	26
Part I - The Stepping Stones Programme - Supporting Teacher's Practices for Autonomy in the Early Primary Classroom	28
Chapter 3 - Methodology to Part I (RQ 1).....	28
3.1 Overview of Chapter 3.....	28
3.2 Methodological Background to Part I - From RCTs to Interpretive Research	28
3.3 Part I Research Questions (R.Q. 1).....	35
3.4 Methods for Part I.....	37
Chapter 4 - Contiguous Analysis: Teachers' Stories of Change Through the Stepping Stones Programme (R.Q. 1).....	52
4.1 Anne's Story	52
4.2 Julie's Story	56
4.3 Lisa's Story	60
4.4 Beth's Story	64
4.5 Conclusion to the Stories of Change.....	68
Chapter 5 - Thematic Analysis: Pockets of Space for Children's Autonomy in an Ecosystem of Teacher Control (R.Q. 1 continued)	70
5.1 Overview of Chapter 5.....	70
5.2 The Teacher as the Centre of Power and Knowledge in the Classroom.....	71
5.3 The Broader Context: The Ecosystem of Teacher Control Beyond the Classroom	80
5.4 Pockets of Space for Children's Choices and Ideas.....	90
5.5 Tensions Between Ways of Working	99

5.6 Finding Commonalities in Successes.....	113
5.7 Alternative Explanations: The Paradigm Shift and the Balance of Methods	119
5.8 Conclusion to Chapter 5.....	123
Chapter 6 - Part I Discussion and Conclusion - Supporting Teachers' Provision of Autonomy and Agency in the Classroom	125
6.1 Developing an Inquiring Stance Towards Changes in Practice.....	125
6.2 A New Set of Skills.....	132
6.3 Further Limitations (Part I).....	135
6.4 Conclusion to Part I	137
Part II - Capturing Inner Motivation: Investigating the Reliability and Variability of the Leuven Involvement Scale.....	139
Chapter 7 - Methodology to Part II (RQ 2).....	139
7.1. Overview of Chapter 7	139
7.2 Methodological Background to Part II	139
7.3 Part II Research Questions (R.Q. 2).....	152
7.4 Methods for Part II.....	159
Chapter 8 - Part II Analyses and Results	168
8.1 Reliability of the LIS: Analytical Approach and Results for R.Q. 2.1	168
8.2 Stability of the LIS: Analytical Approach and Results for R.Q. 2.2.	169
8.3 Variability and Factors Associated With the LIS: Analytical Approach and Results for R.Q. 2.3.....	171
Chapter 9 - Part II Discussion	183

9.1 Suitability of the LIS for the Stepping Stones Project	183
9.2 Engagement as a Characteristic of the Child or the Moment	185
9.3 Activity Setting and Other Contextual Factors Associated with Engagement	186
9.4 Mechanisms for Engagement.....	188
9.5. Further Issues in the Measurement of Engagement	190
9.6 Limitations of the Present Research (Part II).....	194
9.7 Concluding Remarks to Part II	196
Chapter 10 - Overall Conclusion	198
10.1 Overview of Key Findings and Contributions	198
10.2 Synergies Between Parts I and II.....	199
10.3 Implications of This Research for Teachers, Researchers and Policy-Makers.....	203
10.4 Closing Remarks	206
References.....	208
Appendices.....	263
Appendix A - Stepping Stones teacher-researcher Memorandum of Understanding	263
Appendix B - Process for Translating the Theory of Change into Actionable Strategies .	266
Appendix C - Stepping Stones Programme Strategies to Support Children’s Autonomy .	273
Appendix D - Overview of Content for the Stepping Stones Community of Practice Workshops	278
Appendix E - Stepping Stones Interview Protocols.....	285
Appendix F - Running Record Protocol and Template	297
Appendix G - Coding Process.....	306

Appendix H - Validity in Interpretive Research	309
Appendix I - Goals and Action Points Teachers Set Themselves During the Stepping Stones Programme	314
Appendix J - Instruments Reviewed When Selecting a Measure of Inner Motivation (Part II).....	316
Appendix K - Inner Motivation in Learning Piloted Instruments.....	322
Appendix L - Leuven Involvement Scale Pilot.....	328
Appendix M - Description of the Leuven Involvement Scale	335
Appendix N - Teacher-Researcher Memorandum of Understanding for the Leuven Involvement Scale study	338
Appendix O - Opt-In Parental Consent Forms for the Leuven Involvement Scale Study.	341
Appendix P - Activity Setting code descriptions	344
Appendix Q - Leuven Involvement Scale Study Power Analyses	345
Appendix R - Leuven Involvement Scale Study Multilevel Model Statistics	351
Appendix S - Multilevel Models Assumption Checks	353

List of Tables

<i>Table 1 Classroom Characteristics by Teacher in the Stepping Stones Programme.....</i>	<i>39</i>
<i>Table 2 Sample Sizes for the LIS Study</i>	<i>160</i>
<i>Table 3 School Characteristics in the LIS Study</i>	<i>161</i>
<i>Table 4 Fixed Effects of Control Variables in the LIS Study.....</i>	<i>173</i>
<i>Table 5 Intraclass Correlation Coefficients for the Multilevel Models.....</i>	<i>176</i>
<i>Table 6 Predictors of Engagement: Fixed and Random Effects at the School, Class and Child Level.....</i>	<i>178</i>
<i>Table 7 Fixed Effects and Random Effects for the Class-Level Random Slope Models.....</i>	<i>179</i>
<i>Table 8 Fixed Effects and Random Effects for the Model Including Child Characteristics .</i>	<i>181</i>

List of Figures

<i>Figure 1 Controlled and Autonomous Forms of Motivation</i>	5
<i>Figure 2 Framework for the Stepping Stones Project (2016-2019)</i>	9
<i>Figure 3 Place of the Thesis Within the Stepping Stones Project</i>	11
<i>Figure 4 A Theory of Change for Developing Children’s Autonomy in the Classroom</i>	41
<i>Figure 5 Strategies for Choice and Challenge Shared with Teachers</i>	41
<i>Figure 6 Timeline of the 2019 Stepping Stones Programme</i>	43
<i>Figure 7 Children’s Ideas After Looking at a Dead Plant with the Prompt "What Happened?" (Julie’s Class)</i>	97
<i>Figure 8 Visualisations of Contrasting Models of Teacher Control and Children’s Autonomy in the Classroom</i>	122
<i>Figure 9 Mean LIS and Standard Deviation for Each Child</i>	170
<i>Figure 10 Mean LIS and Standard Deviation by Visit Day and by Class</i>	171
<i>Figure 11 Frequency Distribution of LIS</i>	174
<i>Figure 12 Proportion of Observations in Each Activity Setting in Each Class</i>	175
<i>Figure 13 Distribution of Engagement in Different Activity Settings Across All Classrooms</i>	177
<i>Figure 14 Diagnostic Questions for Teachers to Support Agency in the Classroom</i>	205

List of Abbreviations

CI (interview) - Critical Incident (interview)

DfE - Department for Education

EIF - Early Intervention Foundation

OFSTED - Office for Standards in Education

QED - Quasi-Experimental Design

RCT - Randomised Controlled Trial

SEND - Special Educational Needs and Disabilities

SLT - Senior Leadership Team

TA - Teaching Assistant

This thesis is dedicated to my mother, for teaching me the value of childhood, and to my son,
for reminding me of it every day.

Ce que la vieillesse a appris est en fait tout ce qu'elle a oublié.

Romain Gary, *La Promesse de l'Aube*.

Comprendre... Vous n'avez que ce mot-là à la bouche, tous, depuis que je suis toute petite. Il fallait comprendre qu'on ne peut pas toucher à l'eau, à la belle eau fuyante et froide parce que cela mouille les dalles, à la terre parce que cela tache les robes. Il fallait comprendre qu'on ne doit pas manger tout à la fois, donner tout ce qu'on a dans ses poches au mendiant qu'on rencontre, courir, courir dans le vent jusqu'à ce qu'on tombe par terre et boire quand on a chaud et se baigner quand il est trop tôt ou trop tard, mais pas juste quand on en a envie! Comprendre. Toujours comprendre. Moi, je ne veux pas comprendre. Je comprendrai quand je serai vieille.

Jean Anhouilh, *Antigone*.

Preamble - Some Personal Introductory Remarks

I would not be the first – nor the last – to write that education is political. I would argue that so is research. If I have learned anything from this PhD, it is that we come to both (research and teaching) with a baggage of experiences, values and understandings that shape what we think is important, how we go about our work and what we see in the results. All of these come together to shape what we think should be done next.

I came to Cambridge after three years teaching in Primary Schools in inner city London, where my beliefs about education had clashed violently with the practices I was made to adopt. I had not expected the amount of pressure, burnout and battering I experienced, and I was certainly not the teacher I had set out to be. This experience has not just coloured my PhD, it has been its main driver, and the belief that there must be a better way has been the engine to keep me going when I doubted myself.

Then, six months into my PhD, another life-altering event occurred: my son was born. Whilst I owe many of my beliefs about childhood to my own mother, having a child abruptly changed my perspective on early childhood. Seeing my son grow up in parallel with spending hours watching children only a few years older, I could not help but project. I saw him in the five-year-olds made to fill in sheets they did not understand. I saw him in the four-year-olds forced to sit on the carpet for hours. I watched them wriggle and squirm as they struggled to contain their desire to move and act, kept under control by rewards and threats and their desperate desire to please their teacher. I watched a child write only a single sentence in two hours, forced to keep going until it was ‘enough’, and felt how much each letter written seemed to cost him. I watched another sobbing at lunch time, as she had not come back to finish her work as told. All these events were heartbreaking. The very thought of my child going through this was painful.

But I also saw the opposite. I sometimes saw happy and enthusiastic children engrossed in their activities. I saw teachers who cared deeply about the children in their classrooms, about teaching, about making things better. I saw teachers who consoled and cuddled, and bubbled with genuine excitement when a child made a discovery and achieved something new. I saw teachers agonise over details of planning to make sure that every child would be able to benefit, and I saw teachers beat themselves up when they felt they had not met a standard they had set for themselves. Overwhelmingly, the teachers I met wanted what

was best for the children in their care, for their learning and their life chances. Some of the lessons I observed were truly inspiring.

The disconnect between the two – between individuals who are caring and passionate, but who also sometimes end up oppressing children and quashing their love of learning – is not so disconcerting in the light of the findings presented in this thesis. On good days, I find it hopeful. If we can change the system under which schools operate, and if we can support teachers in finding new ways of helping children learn, then school can be for children the place of flourishing that it should be. On bad days, the pressure from the system and the direction reforms are currently taking in England seem unsurmountable. But for me, as my child reaches the brink of Primary school, it is no longer enough to know that there can be another way: we have to make it happen. We have to put children – their wellbeing, their experience of school, and their love of learning – back at the centre of education. This is what I understand by child-centred education. Whilst this thesis is academic in nature, with rational thought and argumentation as its backbone, it is this necessity which is its heart and lifeblood.

Chapter 1 - Introduction

1.1 Thesis Aims

The aim of this thesis is to contribute to our understanding of what teaching practices foster children's inner motivation to learn ('autonomous motivation', Deci & Ryan, 2000) and how we may be able to capture it in order to research it in early Primary classroom contexts (i.e. children aged 4 to 6).

Research suggests that inner motivation is not only an important driver of deep learning and understanding (e.g. Benware & Deci, 1984) but also central to promoting positive attitudes to school and avoiding at-risk students from becoming disenfranchised (Renaud-Dubé et al., 2015; Vallerand et al., 1997). Inner motivation also plays an important role in the development of long-term interests and general personal attributes, such as perseverance, creativity and an affinity towards challenge (e.g. Amabile, 1985; Ryan & Connell, 1989; Tulis & Fulmer, 2013), as well as being important to children's experience of school – their being as well as their becoming (Uprichard, 2008).

However, although most children start school eager to learn, for many children this enthusiasm quickly wanes (Lepper et al., 2005; Nurmi & Aunola, 2005; Spinath & Spinath, 2005). In addition, Primary classrooms in England face enormous challenges in fostering children's inner motivation in school, as accountability measures and pressures from regulatory bodies often result in a focus on test-passing knowledge and controlling teaching styles (Deci, 1992; Mansell, 2007; Valli & Chambliss, 2007). Similarly, research on teaching effectiveness and school improvement has been predominantly concerned with the impact of certain strategies on attainment and grades, which can have a negative impact on inner motivation (Ciani et al., 2010). There is therefore a pressing need to take into account motivation when considering quality in teaching practice.

This thesis is part of a wider project at the PEDAL centre (Play in Education, Development and Learning), the Stepping Stones project, to develop and test a professional development programme for teachers focused on the early years of Primary schooling. The programme aimed to enhance children's inner motivation and executive function skills by helping teachers provide increased opportunities for children to experience autonomy in the classroom. This thesis focuses on the motivation aspect of the Stepping Stones project and is articulated around two axes of inquiry: (1) investigating the changes teachers made to their

practice to increase opportunities for student autonomy through the professional development programme; and (2) searching for a valid instrument to capture inner motivation to be used in large scale testing of this programme at a later stage. These form two distinct parts in the thesis, both taking place within the English educational context.

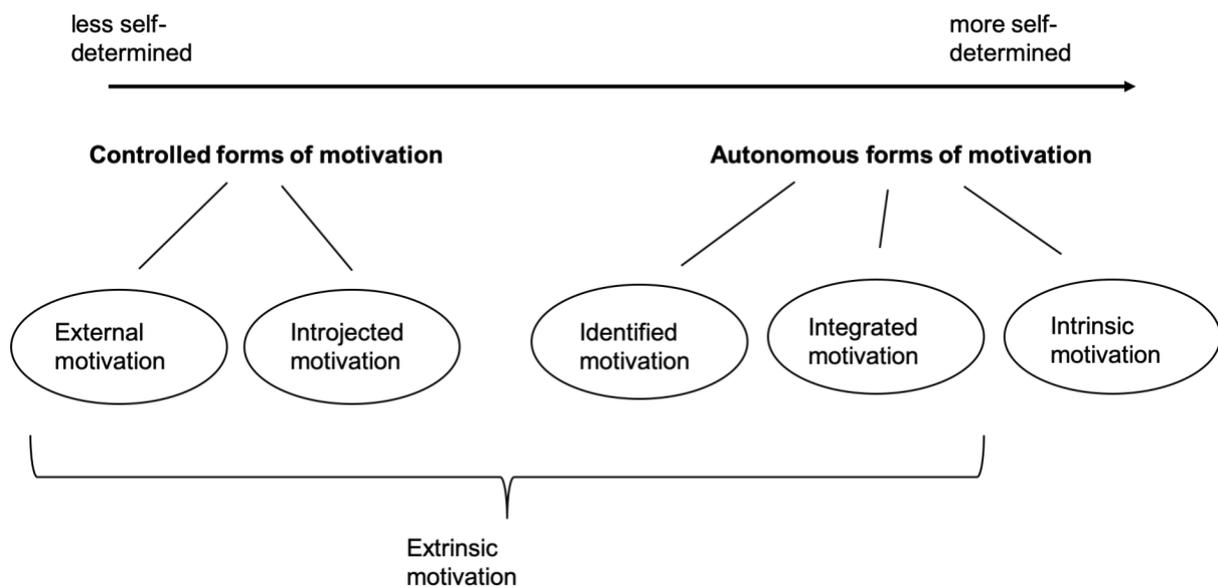
1.2 Theoretical Underpinning: Self-Determination Theory and Autonomous Motivation

This thesis is concerned with children's desire to learn as stemming from internal forces. Wanting to learn, or liking to learn, has been written about in a range of theories. For the most part, this thesis draws from Self-Determination Theory, which describes distinct forms of motivation, that is to say the reasons individuals may have for carrying out a task or action. An action is considered self-determined when it stems from internal rather than external reasons, and different forms of motivation vary in terms of how self-determined they are (Ryan & Connell, 1989; Ryan & Deci, 2000). Intrinsic motivation is understood as wanting to do something for its own sake, for example out of interest or curiosity, or the enjoyment of improving and challenging oneself, whereas for extrinsically motivated actions the reasons for doing a task lie outside the task itself. However, researchers have recognised that tasks may not be carried out for their own sake and still be self-determined (Ryan & Deci, 2000). For example, a task may be tedious but of personal importance to reach a desired goal. The motivation to carry out such a task is very different from the motivation to avoid punishment. Therefore, extrinsic motivation has been further divided into different constructs that vary in the degree to which the reasons for doing a task originate from within the person.

When such reasons emanate from a sense of free will and agency, motivation is thought to be integrated. Identified motivation is similarly self-endorsed, but the reasons may be less fully internalised by the individual (Ryan, 2017). Both integrated and identified forms of motivation are thought to originate from within and therefore are close in nature to intrinsic motivation; these three constructs are grouped together as autonomous motivation as they do not require pressure to be maintained. This is the construct that encapsulates most what is meant in this thesis by the inner desire to learn. However, the term 'autonomous' is both unwieldy and unintuitive for those unfamiliar with the theory, as well as repetitive and possibly confusing in a field that is concerned with children's autonomy. Therefore, throughout the remainder of this thesis I use the lay term 'inner motivation' to refer to the construct of 'autonomous motivation' as understood in Self-Determination Theory.

Inner motivation is distinct from less self-determined forms of motivation, called controlled motivation (Figure 1). Controlled motivation encompasses two constructs where the locus of control lies outside the person, that is to say that a person acts not from their own desire but because of forces putting pressure on them. Introjected motivation relates to reasons for doing the task that result from internal pressures such as feelings of guilt and anxiety (Deci et al., 1994). External motivation results from external pressures; these can be direct, in the form of rewards, punishments and threats, or more subtle, for example through expressions of approval or disapproval by others. Different forms of motivation are also conceptualised in opposition to amotivation, the absence of motivation.

Figure 1 *Controlled and Autonomous Forms of Motivation*



Note. Adapted from Ryan & Deci's (2000) quasi-simplex model of motivation.

1.3 Rationale for the Thesis

1.3.1 The Importance of Inner Motivation

It is the position taken here that school ought to be a place where children thrive on learning, rather than one where learning is imposed onto children and which children engage in out of compliance or coercion. But if children's inner motivation ought to be valued as an important goal in itself (Ryan, 2017), inner motivation also has a large range of educational benefits.

Inner Motivation and Learning Outcomes. A large body of research shows that students who are motivated from within tend to learn more and at a deeper level (e.g. Benware & Deci, 1984; Grolnick & Ryan, 1987; Jones et al., 2015; Ryan et al., 1990). Evidence suggests that children who have higher intrinsic motivation in academic subjects are also more likely to be successful in school (Becker et al., 2010; Gottfried, 1985; Gottfried et al., 2013; Taylor et al., 2014) and inner motivation is key to achievement (Fortier et al., 1995; Guay et al., 2010; Lemos & Veríssimo, 2014). In particular, interest and curiosity are related to improved learning (Anmarkrud & Bråten, 2009; Arnone et al., 1994; Dunst et al., 2011; O’Keefe & Linnenbrink-Garcia, 2014; Rotgans & Schmidt, 2011) and better memory (Gruber et al., 2014; Guthrie et al., 2007; Kang et al., 2009; Schiefele & Krapp, 1996). Indeed, the importance for learning of being curious has led some researchers to call it one of the pillars of achievement (Stumm et al., 2011).

By contrast, controlled motivation tends to be related to poorer outcomes. Although a focus on grades can be beneficial to performance on tests (e.g. Elliot & Church, 1997; Harackiewicz et al., 1997), improved test scores do not always equate better learning (Ryan et al., 1990). For example, in one study students who had an intrinsically motivating purpose for learning materials had significantly higher conceptual learning scores than students who were told to study for a test, though both groups fared equally well on rote learning tests (Benware & Deci, 1984). In addition, in a longitudinal study with elementary school children, children who read for external reasons such as parental pressure had poorer reading skills than children with lower extrinsic motivation, even when controlling for past achievement and for how much they read (Becker et al., 2010). Indeed, children who do not like reading are unlikely to create for themselves the opportunities to engage with books and to improve their reading skills (Guthrie et al., 2007).

Positive Long-Term Attitudes to School. Research suggests that inner motivation is important for building positive attitudes to school and future success. Students’ intrinsic motivation predicts their intention to stay in school (Renaud-Dubé et al., 2015), whereas students who do not feel self-determined are more likely to drop out of school (Vallerand et al., 1997). Halting disengagement is crucial to dealing with underachievement and disaffection with the school system (Duffy & Elwood, 2013; Finn & Rock, 1997). For example, one group of researchers found that introducing intrinsically motivating activities such as gardening led to a substantial decrease in school failure, school dropout, and disruptive behaviours (Ruiz-Gallardo et al., 2013). Engaging students with learning is also key to their developing future

interests, as positive experiences in a particular domain are thought to lead to longer-term interests (Deci, 1992; Mitchell, 1993; Renninger et al., 1992).

Other Benefits. Inner motivation has benefits over and above school achievement. Inner motivation is thought to be related to greater well-being (Levesque et al., 2004) and happiness and vitality (Nix et al., 1999). By contrast, students with introjected forms of motivation are more likely to cope badly with failure and to experience anxiety (Ryan & Connell, 1989; Vallerand & Bissonnette, 1992).

Inner motivation is also related to a range of behaviours that help learning such as persistence, effort, and self-regulation (Harter, 1978; Jones et al., 2015; Lauriola et al., 2015; O’Keefe & Linnenbrink-Garcia, 2014; Ryan & Connell, 1989; Tulis & Fulmer, 2013). By contrast, extrinsic forms of motivation such as a focus on grades has been associated with negative learning behaviours such as self-handicapping and disruption (Midgley et al., 2001).

Finally, intrinsic motivation is also associated with creativity. For example, when students are put in a situation that creates extrinsic reasons for engaging with it, they also tend to become less creative (Amabile, 1985).

1.3.2 The Decline of Motivation in School

Most children begin school with positive attitudes and beliefs (Dockett & Perry, 1999; Freedman-Doan et al., 2000). Young children are curious about the world around them and ask huge numbers of questions in order to learn and understand (Chouinard et al., 2007; Tizard & Hughes, 1984). Children are also intrinsically moved to master new skills (Carlton & Winsler, 1998). All this they do without the need for adults to prompt or reward them. Yet many studies have shown that as children go through school, this intrinsic force for learning steadily declines (e.g. Anderman & Maehr, 1994; Gottfried et al., 2001; Lepper et al., 2005; Nurmi & Aunola, 2005; Otis et al., 2005; Spinath & Spinath, 2005). It could be that young children are simply more curious and intrinsically motivated than older individuals. However, a number of studies suggest that this is not the whole story, and that something happens in schools that affects students’ intrinsic motivation.

First, there is evidence that children’s intrinsic motivation does not altogether disappear. Children of different ages show similar levels of exploratory behaviour, a sign of curiosity, when presented with a curious object (Engel, 2011; Jirout & Klahr, 2012). Indeed,

even adults can be motivated to learn for its own sake when given the chance (Abuhamdeh & Csikszentmihalyi, 2012a; Csikszentmihalyi, 1990).

Secondly, students' interests and curiosity can be fostered or lowered by parents and educators (Baroody & Diamond, 2016; Gottfried et al., 2016; Patall et al., 2018; Schukajlow & Rakoczy, 2016; Shernoff et al., 2003) and it seems that the decline in motivation is more pronounced at school than in other settings. For example, in one study the number of questions that children asked plummeted from an average of 26 questions per hour in the home to only two in the classroom (Tizard & Hughes, 1984). One researcher noted that one of her studies had to be modified because children's expressions of curiosity in classrooms were so rare (Engel, 2009).

Finally, school tends to be related to constructs that negatively influence students' motivation. One study found that two thirds of US high school students were bored at least every day (Yazzie-Mintz, 2010) and boredom and apathy have been suggested as the main reasons for student's lack of engagement in school (Pekrun et al., 2010).

There is therefore a pressing need to help teachers adopt practices that foster rather than extinguish children's inner motivational resources.

1.4 Situating the Thesis

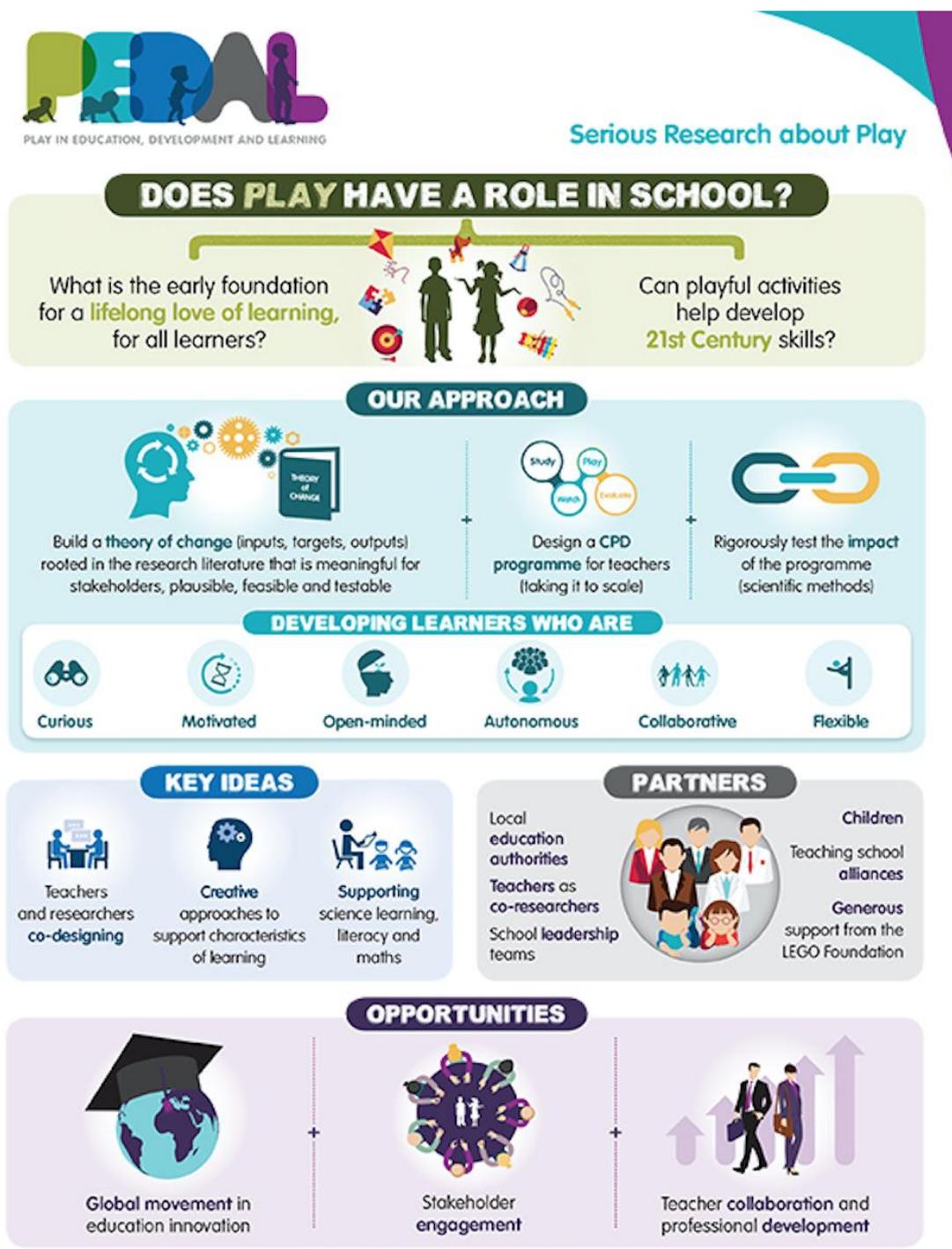
1.4.1 The Stepping Stones Project

This thesis forms part of a larger project, the Stepping Stones project, which targeted children's inner motivation to learn as well as other goals. The Stepping Stones project was concerned with developing a professional development programme to help teachers support students' 21st Century skills by considering the role of playful activities in schools (Figure 2). This work has gone through different cycles of refinement, and it is considered to be under constant improvement from each iteration. In the initial stages, the team built theories of change that focused on curiosity, motivation and problem solving, and worked with a small group of Reception and Year 1 teachers (i.e. of children aged 4 to 6) to investigate learner-centred teaching (Kittredge et al., 2018).

At the time the present thesis was in development, the team was focused on children's inner motivation and executive functions. We designed a theory of change to clarify the links between these psychological mechanism and children's autonomy and the strength of the evidence for different practices to promote these outcomes. My own review of the literature contributed to these discussions by providing evidence on inner motivation. This work

culminated in a professional development programme which aimed to encourage teachers to experiment with their practice in order to promote children’s autonomy in the classroom. We trialled this programme through a Community of Practice with a group of teachers in England in the spring of 2019, and Part I of this thesis is concerned with understanding what changes teachers were able to make, as well as the barriers and affordances they faced.

Figure 2 Framework for the Stepping Stones Project (2016-2019)

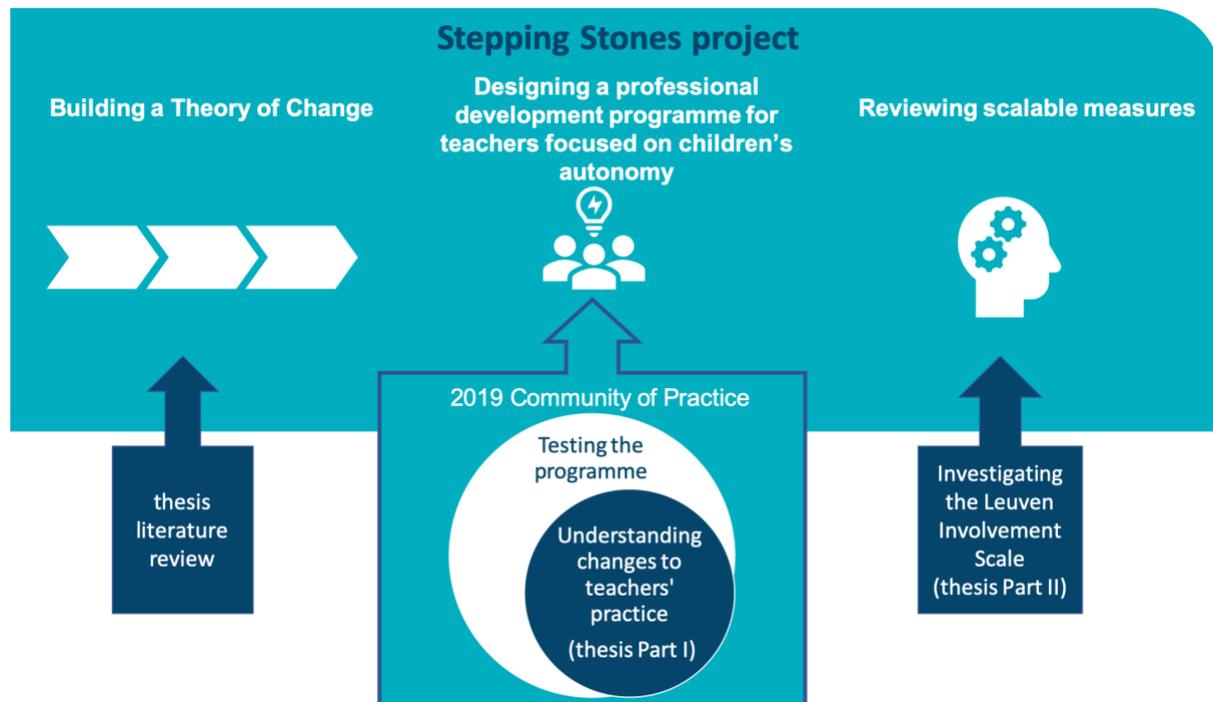


At the same time as developing the programme, the Stepping Stones team, including myself, was concerned with how we might provide evidence for the effectiveness of the programme, and we met with various stakeholders including the Early Intervention Foundation (EIF). At the time of project development (2018-2019), the aim was for the project to eventually lead to a full Randomised Controlled Trial (RCT) or an appropriate equivalent for schools. Therefore, we were interested in finding instruments that would allow us to capture the effectiveness of the programme at scale not only on learning, but more specifically on the constructs we were targeting, including inner motivation. Therefore, part of my thesis is also concerned with finding an appropriate instrument to be used in further trials of the Stepping Stones programme in the age group we were targeting, i.e. the start of Primary school. After reviewing a number of instruments, I focused my research on a promising measure of young children's deep engagement in activities, the Leuven Involvement Scale (LIS, Laevers, 1994). My contribution to the project through this thesis was therefore threefold, as can be seen in Figure 3:

- the literature review contributed to the Theory of Change and programme development, specifically relating to inner motivation (summarised in Chapter 2), whilst other team members worked on self-regulation.
- Part I (Chapters 3, 4, 5 and 6) contributes to our understanding of the Community of Practice programme of 2019 and how teachers can change their practice in order to increase opportunities for children's autonomy in the classroom.
- Part II (Chapters 7, 8 and 9) contributes to our search for an appropriate instrument to measure inner motivation in young children by studying the validity and reliability of the LIS in research and evaluation contexts.

In the conclusion, I outline the main points and contributions this thesis makes. It should be noted that the project is no longer called 'Stepping Stones', as it has evolved along with our thinking. However, because that is what it was called at the time, I continue to refer to it as such throughout this thesis.

Figure 3 *Place of the Thesis Within the Stepping Stones Project*



1.4.2 Positionality

My own positionality in this research is complex. Before coming to Cambridge, I had trained then worked as a Primary school teacher in London. I had been a ‘struggling teacher’ (Culshaw, 2019), in particular as my own vision for education conflicted with practices that had been imposed on me, and I often worried about the negative impact these practices might be having on children. My main interest in coming back to academic studies was in fostering children’s “love of learning” (Le Courtois, 2017) and on many levels this PhD felt highly personal. My experience as a teacher also meant that I had insider knowledge and could understand ‘teacher speak’. I was familiar with the daily rhythm of schools and was comfortable in classrooms. However, this position also was in tension with my role as a researcher. Teachers interacted with me in ways that suggested that, to them, I was an outsider, albeit one who understood their world, and this created some distance in the relationship – mutual respect but not fellowship. In addition, whilst this was intentional because of the nature of the research, my position was one of observer. This dual position – outwardly and functionally an outsider but internally experiencing the research as insider – sometimes resulted in tensions: between neutrality and subjectivity, between observing and intervening, between describing and interpreting. I explore some of these issues further in

Chapter 3, but many were instead grappled with outside of these pages, and some have possibly remained unresolved.

In addition, I am also a mother, and my child, who was born early on during this PhD, has grown in step with this thesis. This has not only coloured my understanding of young children by providing me with an additional lived experience – albeit a different one from teaching – it also affected my observations emotionally. It deepened the empathy I felt for the children in the classrooms I visited. As I describe in Chapter 7, empathy is part of the way the Leuven Involvement Scale is used, and in some ways this increased attunement to children may have helped my understanding and use of the instrument. However, at the same time, it requires caution, as I could easily project my own fears and sense of oppression onto children who did not share these feelings. Therefore, I have tried to be cautious in moving from observations of teacher control to inferences about children’s experience of it, whilst considering the ways in which children demonstrated signs that their need for autonomy was not met or that their inner motivation was not supported.

1.4.3 A Thesis at Crossroads

Because children grow up, we think a child's purpose is to grow up. But a child's purpose is to be a child. (Stoppard, 2013, p. 216)

This thesis is situated at an uncomfortable crossroads between opposing paradigms. The research took place within a larger research project which aimed to develop a professional development programme for teachers, with specific targets and outcomes, ultimately leading to testing the programme with experimental or quasi-experimental methods. As I argue in Chapter 3, the nature of this undertaking, as well as the language associated with it, are clearly positivist. Not only this, but they fit within the current rhetoric both in policy and in some education research of experimental work as being the ‘gold standard’ for causal inference and more broadly within a paradigm where the goal of education is the production of outcomes. Yet a large part of this thesis (in particular Part I) is anchored in qualitative research, where the aim is not to produce ‘proof’ that the programme ‘works’, but to advance our understanding of what happens in classrooms. Not only this, but the heart of this project is not in outcomes – it is a vision of childhood and learning as portrayed in the above epigraph, far from target-setting and instead about nurturing rather

than shaping children (Gopnik, 2016). At the same time, Part II of this thesis focuses on a measure of inner motivation as a possible outcome for the Stepping Stones project.

Such a position may appear contradictory, and it is indeed an uncomfortable one, which requires constant re-evaluation of assumptions. But I argue below that it is not an incoherent position, and that it is one of value – indeed it may provide real solutions to the problems described above.

Reciprocity of Different Epistemologies. All research methods have limitations, not only in terms of the *kinds* of knowledge they can produce, but also in terms of how much of that knowledge they can claim to have produced. Being at a crossroads, those limitations are particularly apparent in this thesis, and discussions of those limitations form an important part of it. But the answer to the limitations of any methodology should not be to dismiss it entirely, nor to pretend the limitations do not exist. If different methodologies give us different facets of reality, then by combining them we can begin to build a better picture. In the current project, qualitative research contributes to our understanding of how specific teaching practices give rise to different power structures and experiences for the children and how different constructs are enacted in the classroom. However, gaining recognition that the programme or specific practices have consistent effects requires engaging in experimental designs, whilst acknowledging their limitations. Taking the project to scale thus requires measurement and considerations of validity. The aim here is not to resolve these tensions or provide a final verdict on which way is best, but to acknowledge that this thesis is firmly ensconced in these issues, which recur throughout.

Reconciling Views of Childhood. I have described that this thesis is concerned with children's authentic experiences and inner motivation as intrinsically valuable, whilst also arguing for their contribution to educational outcomes; in addition, a large part of this thesis is dedicated to a measurement that reduces children's experiences to a rating. However, rather than creating a contradiction, viewing childhood and children's experiences as valuable in and of themselves, without reference to adult-determined goals, enriches the discussion around what constructs such as motivation or autonomy might mean in practice, and keeps in check tendencies to reduce children to data points, even when research is carried out quantitatively. Yet it is a fine line between wanting to provide the right environment for children to flourish, and having 'flourishing' (whichever way that is conceptualised) as a target, or worse, an expectation. Here perhaps lies the strongest tension: between the current

moment and the future, and around how much children's freedom ought to be restricted in the name of goals adults have set. Whilst this thesis does not directly address those questions, they are part of the underlying web of issues that surround it. These tensions result from broader questions around the meaning and value of autonomy, and around the tension and reciprocity between 'being' and 'becoming' (Uprichard, 2008), which have played an important role in shaping this research and my own thinking.

1.5 Summary of Research Aims

This thesis is part of a broader project, the Stepping Stones project, which was concerned with fostering children's 21st Century skills and their lifelong love of learning, with a focus on inner motivation and executive functions through increasing children's autonomy in the classroom. After outlining the literature and Theory of Change that underpins the programme (Chapter 2), this thesis splits into two main parts, with different research aims and methodologies.

Part I focuses on the professional development programme created through the Stepping Stones project, which encouraged teachers to increase opportunities for children's autonomy in the classroom. Specifically, in this thesis I explore and explain what changes teachers were able to make through taking part in the Community of Practice we ran in Spring 2019 in England, using interpretive methods.

Part II of this thesis focuses on a promising instrument that could be used in future iterations of the Stepping Stones project to quantify children's quality of engagement in classroom activities. The aim of this research is to investigate the suitability and validity of the instrument for these purposes.

Chapter 2 - Review of the Literature on Fostering Children's Inner Motivation by Supporting Their Autonomy in the Classroom

This section presents an overview of the literature that supported the part of our Theory of Change concerned with children's motivation. Specifically, I present evidence suggesting that by supporting children's autonomy in the classroom, teachers would foster greater inner motivation to learn as well as how this might be translated into classroom practice.

2.1 Autonomy as Supportive of Inner Motivation

2.1.1 Motivational Needs in Self-Determination Theory

Self-Determination Theory predicts that individuals have a number of needs which need to be met for them to experience inner motivation (Ryan & Deci, 2000). They are:

- (1) the need for autonomy, i.e. a sense of choice and control over the work to be done;
 - (2) the need for competence, which relates to feeling that one is able to do the task well;
- and
- (3) the need for relatedness, which includes positive relationships and a sense of belonging and acceptance.

A large body of research supports the importance of these needs for motivation (see Ryan, 2017 for a review). This thesis focuses on the need for autonomy, which is considered the most important to inner motivation (Ryan & Deci, 2000) and which is undermined by controlling practices in schools (Assor et al., 2005; Flink et al., 1990). It should be noted that autonomy refers here not to the action of doing something independently, but to the internal state experienced by someone, and in fact autonomy can be experienced in any act, whether individual or collective (Ryan & Deci, 2000). Autonomy is more broadly understood as "the need to regulate one's own behaviour and to govern the initiation and direction of one's actions" (Carlton & Winsler, 1998, p. 160). When this need is met, actions are experienced as originating from within one's self, and individuals feel in control of and ownership over these actions. By contrast, when this need is thwarted, individuals feel as though they are being controlled or coerced by others and therefore feel little responsibility for their actions (Carlton & Winsler, 1998). The feeling of autonomy can be further decomposed into three distinct constructs: perceived choice, volition and internal locus of causality (Reeve et al., 2003). Perceived choice refers to whether or not someone feels they were offered a choice; volition refers to how free rather than forced a person feels they are when engaging in an action; and

locus of causality refers to an individual's perception that their actions are initiated and controlled by themselves rather than by external forces (Reeve et al., 2003). Deci (1992) argues that children fail to develop interests in activities that take place in contexts that thwart their motivational needs. Consequently, research has looked into teacher practices that support motivational needs, and in particular the need for autonomy. This focus on autonomy should not be taken as a disregard for the importance of the other two needs in Self-Determination Theory. However, our experience working with teachers, as well as my own working in schools and existing literature, pointed to autonomy as the most pressing need to support and the one most clearly thwarted in classroom environments. In addition, we felt that teachers, in particular of young children, were already aware of the need to support children's feelings of competence and belonging, and were particularly proficient at creating nurturing, supportive relationships. Therefore, our programme targeted autonomy and autonomy-supportive teaching.

2.1.2 Autonomy-Supportive Teaching

What is Autonomy-Supportive Teaching? Autonomy-supportive teaching is about helping students connect their learning with their sense of self, and this can take different forms. One study operationalised autonomy as the extent to which actions fitted with students' own interests or goals and sense of choice, and found that meeting this need meant students were more likely to be happy about their learning, and to be interested in it (Flunger et al., 2013). In another study, autonomy-support was described as teachers providing relevance and choice, and allowing criticism (Assor et al., 2002). In Reeve and Jang's study (2006), autonomy-supportive teachers were those who spent time listening to students, allowed students to work in their own way, and provided time for students to talk. Generally, autonomy-supportive practices require the teacher to provide opportunities for choice, to provide rationales when choice cannot be given, and to be sensitive to students' perspectives. Controlling practices, by contrast, include frequent directives, controlling the pace of learning and not allowing critical and independent opinions. Controlling behaviours can also be less explicit, for example by offering or withdrawing material or emotional rewards (e.g. Assor et al., 2004) or making performance goals particularly salient (e.g. Midgley et al., 2001).

Benefits of Autonomy Support. Evidence suggests that teachers who support their students' need for autonomy also support their students' intrinsic motivation (Boggiano et al., 1988; Reeve et al., 2004) and enjoyment of learning (Walsh et al., 2006).

By contrast, controlling practices are associated with less inner motivation. By being overly controlling, teachers might disempower children, who may lose the initiative for their learning (Flink et al., 1990; Grolnick & Ryan, 1987; de Kruif et al., 2000). Children are also less likely to be engaged when teachers are being directive (Almqvist & Almqvist, 2015; Powell et al., 2008). Performance goals similarly can also erode motivation by creating classroom cultures focussed on outwardly demonstrating competence rather than on self-improvement (Ciani et al., 2010). Finally, more controlling forms of teaching lead to children being less curious and exploring less (Engel, 2011; Murray, 2012).

2.1.3 Choice: Opportunities for Autonomy and Interest

The Central Role of Choice in Supporting Inner Motivation. Provision of choice and participation in decision making is central to the notion of autonomy support. Despite this, development programmes for teachers derived from Self-Determination Theory tend to focus on other aspects of autonomy support rather than emphasise choice per se (e.g. Cheon et al., 2016; Jang et al., 2016; Reeve et al., 2004). Yet studies looking explicitly at choice have shown that it is an important driver of inner motivation (Patall et al., 2008). For example, pre-school children are more likely to be positively engaged when activities provide a greater degree of choice such as in 'free flow' settings (Vitiello et al., 2012). Choice also plays an important role in supporting deep learning and self-regulation (Kuhn & Ho, 1980; Perry, 2013). Finally, children seem to benefit more from choice than adults do, possibly because they have fewer opportunities to exercise their agency (Patall et al., 2008). However, there are also conflicting results regarding the effects of choices on motivation which need to be further discussed.

The Choice Conundrum. A number of experimental studies have shown that providing choice does not necessarily lead to positive learning or motivational outcomes, and can even sometimes have negative impacts (Iyengar & Lepper, 2000). For example, d'Ailly (2004) found that students' interest level was not affected by provision of choice whilst Flowerday and colleagues found that it is interest rather than choice that predicts depth of learning (Flowerday et al., 2004; Flowerday & Shell, 2015).

According to a meta-analysis (Patall et al., 2008), choice can be beneficial under certain conditions, but detrimental in others. Students with low self-efficacy may engage less with work when given choices because of the additional uncertainty brought by the choices (Flowerday et al., 2004). Students may reject choice or resent being given choices if it increases their performance anxiety (Clifford, 1991) or because they expect authority figures to make such decisions (Iyengar & Lepper, 1999). If individuals are not given enough information to make a sound judgment, individuals might disengage and choose arbitrarily (Schwartz, 2000). Finally, making choices is effortful and individuals' cognitive performance can decrease because of fatigue (Baumeister et al., 1998).

Defining Choice. Part of the issue is one of confusion over what choice refers to in different situations. Choice in the classroom is generally understood as providing opportunities for students to make decisions about their learning, but there are many ways of categorising different types of choices. Reeve et al. (2003) differentiate between option choices and action choices. Option choices require individuals to choose between two or more options, which may not increase motivation or engagement, whereas action choices involve self-regulation of the learning activity. Similarly, Katz & Assor (2006) distinguish two forms of choice, "picking" and "choosing". They define choosing as allowing one's preferences to be expressed, which supports feelings of autonomy, while picking only entails a choice amongst limited options.

In classrooms, choice can also be understood as a participatory process where the direction of the learning is co-decided by the teacher and the students. Zahorik (1996) found that teachers develop students' ownership of classroom events by involving them in planning units and in choosing tasks. For example, students might participate in decisions around which topics to explore, which texts to read, the sequence of texts, and the particular skills to emphasise.

Clarifying the Effect of Choice in Experimental Studies. What may explain the ambiguous effect of choice in experimental studies is that these situations always offer choice between a limited number of options. There is evidence that such choices may increase perceived choice, but not feelings of volition and internal locus of causality. Specifically, Reeve et al. (2003) showed that intrinsic motivation was increased when choice was designed to enhance the perception of an internal locus of causality and volition, by giving individuals the freedom to initiate and regulate their own behaviour (action choices). By contrast, providing options amongst tasks determined by the experimenter only increased perceived

choice but not intrinsic motivation. Indeed, presenting options in an experimental setting may even be perceived as a controlling situation (Patall et al., 2008; Moller et al., 2006).

Another explanation for the mixed findings regarding the effect of choice is that experimental studies often attempt to separate the effect of choice from the effect of interest. It may be that choice increases inner motivation through interest rather than through perceived choice. For example, Assor et al. (2002) found that choice promoted motivation when it allowed students to make selections that were relevant to their individual interests. Therefore, attempts to disentangle the two results in choices that are not opportunities to pursue interests, and therefore meaningless. In particular, in the studies by Flowerday and colleagues, participants were asked choose between packets simply labelled A or B, with no other information (e.g. Flowerday & Shell, 2015). In another study, Wilde et al. (2018) found that matching students with a preferred, meaningful option increased motivation more than asking students to choose between those two options. It may be that being given an opportunity to engage with a preferred option increases feelings of volition and internal locus of causality, even if it does not explicitly involve picking the option. As Lee & Hannafin (2016) argue, “providing only a single option may improve perceived autonomy when one truly endorses that option” (p. 718).

Summary on the Role of Choice in Autonomy Support. In our Theory of Change, choice is understood as opportunities for students to become involved in the decisions regarding their learning, so that the learning is personally relevant, students have opportunities to pursue questions and interests, and students experience a feeling of autonomy. It implies a sharing of power (Almqvist & Almqvist, 2015),

Nonetheless, experimental studies on choice provide a number of important lessons for providing choice in the classroom. First, meaningless choices, where individuals do not know what they are choosing or care little about the outcome, have no or negative impacts on motivation and learning. Secondly, choices seem to be important insofar as they allow students to pursue what interests them. Finally, choices can be effortful, and this can be taxing for students’ cognitive resources. However, this is only an issue if one assumes choices are always an individual activity, made unguided.

Choice and Structure, not Choice or Structure. Providing choice and autonomy is about empowering students to take control of their learning. But without the guidance of their teacher and without a structured learning environment, students may be confused about the

goal of the learning, the steps needed to achieve it, or may access tasks that are too easy or too difficult for them. As previously described, a feeling of competence is a key element of feeling self-determined, and therefore providing choice should not be at its expense.

A number of studies show that supporting both children's learning and their motivation requires a balance of autonomy and structure (Guay et al., 2017; Jang, Reeve, & Deci, 2010; Perry, 2013; Skinner & Belmont, 1993; Vansteenkiste et al., 2012). Structure refers to those teaching strategies that guide students' activities, such as providing clear instructions and expectations, highlighting meaningful learning goals, and supporting students' endorsement of classroom rules and routines (Jang et al., 2010). A number of studies have found that in classrooms high in both autonomy and structure, students are more motivated and are better able to self-regulate (Perry, 2013; Sierens et al., 2009; Skinner & Belmont, 1993; Vansteenkiste et al., 2012). Structure is thought to help students feel empowered by making the learning environment consistent and predictable and by enhancing perceived competence (Jang et al., 2010), in particular when teachers provide tasks and materials at an optimal level of challenge (Guay et al., 2017).

2.1.4 Self-Determination Theory and Young Children

Whilst a large body of evidence supports Self-Determination Theory, much of it comes from research with older students. However, this thesis is focused on children at the start of Primary school (aged 4 to 6). Given the importance developmental differences across age groups, there is a risk that the importance of autonomy for younger children – who are still very dependent on adults – may have been overstated. However, below I outline two main reasons why autonomy is in fact hugely relevant to working with young children.

First, some of the research from Self-Determination Theory has been carried out with Primary-aged children, for example on the use of rewards (Lepper & Greene, 1975), informational versus directive language (Amabile, 1985), or controlling teaching (Flink et al., 1990), so the theory has been applied and tested with children. In addition, research from other fields also emphasises the importance of autonomy for young children (e.g. Forman, 2007; Graves & Larkin, 2008; Green, 2013; Perry, 2013). For example, parental autonomy support refers to parental behaviours that support children's goals, interests, and choices and has been shown to support infant exploration (Hughes et al., 2018).

Finally, Self-Determination Theory conceptualises autonomy not as a condition of independence, but as a need to experience situations as stemming from volitional sources. It is

therefore likely that the way in which the experience of autonomy is supported in children will be different from how it is supported in older students and in adults. For example, the importance of choice for intrinsic motivation has been shown to be more important for children than in older age groups (Patall et al., 2008). We could also suppose, for example, that it is possible that language considered controlling by teenagers may not be experienced as controlling by young children, or that different contexts (e.g. school versus home) lend themselves to different interpretations of the same actions. Indeed, there may be variations in whether actions are supportive of inner motivation because of how they are interpreted differently not only across age groups, but also across cultures (Lee et al., 2018), and likely across individuals and across situations.

2.2 Autonomy and Choice in Child-Centred Pedagogies

So far, this literature review has focused on evidence from academic research, with an emphasis on experimental studies. However, children's autonomy in learning and participation in decisions about their learning have long been concerns for educators (Chung & Walsh, 2000; Lillard, 2019). Many practices that are supportive of autonomy have therefore been developed on the ground by practitioners before later being studied by researchers. Here, I draw on these pedagogies to explore how children's autonomy might be supported in practice.

2.2.1 Learning Through Play and Playful Learning

Whilst there is no consensus on what exactly defines play, one of its agreed characteristics is that individuals willingly engage in it. Individuals cannot be forced to play or be playful, or it would not be play anymore. As such, both autonomy and inner motivation are central to the concept of play. In addition, in academic settings where children have many opportunities for play, they also often have a large amount of autonomy as to what activities they do at any given point in time.

Researchers have argued that play has a number of educational benefits, including the development of language, social-emotional skills and self-regulation (Singer et al., 2006; Whitebread et al., 2012), as well as being important for children's mental health and wellbeing (Whitebread, 2018). Not only this, but some have argued that often play and learning activities are not distinct in the Early Years, in particular when children are involved in explorations and inquiries of their own initiative (Samuelsson & Carlsson, 2008). This has

led to a focus on researching the role that play can have in academic settings. In Early Years settings, where teachers must meet curricular goals as well as provide opportunities for play, educators argue that adults can use children's play as opportunities for learning whilst following children's interests and initiative. For example, one popular approach in England is "planning in the moment" (Ephraive, 2018), where adults seize learning opportunities in children's play by supporting children's next steps in learning as and when children show interest and curiosity in something. However, in order to maintain children's sense of autonomy and play, this requires adults to follow the child's interests rather than "hijack" their play (Gooch, 2008, p. 95) for their own purposes.

In addition, educators and researchers have argued that the characteristics of play (e.g. that it is meaningful, iterative, active and joyful, Parker & Thomsen, 2019) can be harnessed to meet learning objectives in more structured activities such as inquiries and 'guided play' (Skene et al., 20022; Weisberg et al., 2013). For example, in one study young children learned about the properties of geometric shapes through one-to-one sessions with a researcher framed as a game, where the children had to discover the 'secret' of each shape (Fisher et al., 2013). They had the opportunity to explore the shapes to uncover the 'secrets', but were also guided by categorised exemplars, questioning from the adult, encouragements to manipulate the shapes, and an activity to apply their new knowledge. The study found learning benefits to the guided play approach: children who had received a guided play intervention performed better on later tests of their understanding of shapes than children who had only had an unguided free exploration of the shapes, and those who had received formal instruction from the researcher.

In order to distinguish between free play, where children are in complete control over what they do and learn, and more adult-led playful learning activities, some researchers have framed these differences in terms of a continuum or spectrum of play (Jensen et al., 2019; Pyle & Danniels, 2017; Zosh et al., 2018). The amount of control children have over the activity – which relates to their autonomy – is used as the basis for this continuum. The more control children have to initiate and direct the activity, the more it is considered like free play. At the opposite end of the continuum, adults initiate and direct the activity in order to meet specific learning goals. Whilst this conceptualisation may be helpful to clarify different forms of play, there is also a risk that the place of children's autonomy and inner motivation is lost through a focus on the instrumental value of play for learning (Wood, 2019). In other words, whilst these adult-led playful activities may retain some of the characteristics of play, without an explicit focus on how autonomy is supported, children's inner motivation may not

be fostered. For example, in the guided play example above, whilst the activity was playful and the children had some autonomy as to how they explored the shapes, the task was also highly structured and had a definite learning point. The activity was designed to have game-like attributes, rather than considering opportunities for children to experience the activity as their own and stemming from their own will.

2.2.2 Autonomy in Alternative Pedagogies

Children's autonomy is also emphasised in a number of alternative pedagogies, that is to say which operate outside of the conventional models of schooling. Whilst there are many such models, here I briefly review how autonomy relates to two influential pedagogical systems which provide children with autonomy in different ways: Montessori and Reggio Emilia. Whilst there are many other alternative education models (e.g. Freinet, Steiner-Waldorf, Sudbury, democratic schools), Montessori and Reggio Emilia are both focused on the Early Years (up to 6 years of age) and well-researched models of education. They are also models which are drawn on throughout this thesis and therefore it is useful to introduce them here.

Montessori Pedagogy. Montessori is one of the most widespread alternative pedagogies in the Western world (Lillard, 2019). Based on a view of the child as capable and eager to learn, Montessori pedagogy relies on a structured, prepared environment where children can access clearly organised activities of different levels of difficulty. Montessori Early Years education emphasises children's inherent curiosity to learn and desire to master specific skills. Children's autonomy and their inner motivation to learn are central to Montessori education. Children have large amounts of freedom within a structured environment, i.e. freedom within boundaries (Lillard, 2019). Children do not follow a timetable and strict edicts set by the teacher, and instead are free to choose what activities to do and for how long and with whom. Unlike other Early Years settings, however, the focus of the activities is not play but the development of specific competences children are intrinsically motivated to master (Lillard, 2013).

The teacher's role is to prepare an environment that will allow children to engage in stimulating activities and to observe the child in order to understand how their interests and their needs develop on an individual basis. This paves the way for teachers to introduce specific activities and materials that children might be ready for at opportune times, and to

show the children how to use them. Children also learn by watching peers engaged in activities and by playing together. Structure in the physical environment is thought of as another teacher, and materials are freely accessible on open shelves, arranged in an order that is logical for the child, such as from least to most challenging. Children's concentration is protected from distractions by the use of work mats and being free from interruption. Therefore children are supported in maintaining their focus on tasks that matter to them and to meet goals they have set themselves.

Reggio Emilia pedagogy. Like Montessori, Reggio Emilia is characterised by a strong emphasis on the natural capacities of the child to be curious and to seek knowledge (McNally & Slutsky, 2017). The Reggio Emilia approach also focuses on the myriad ways in which children express themselves, including non-verbal forms of communication like painting and dramatic play. Connected with this idea, there is a strong emphasis placed on learning activities within the social and cultural environment of the child. Children's autonomy in Reggio education resides in their place as leaders of their own learning, with the teacher acting as a stimulating and supporting guide rather than as a directing and coercing force (McNally & Slutsky, 2017). Kim and Darling (2009) give the example of a group of 4-year-old children studying a Monet painting in a Reggio-inspired classroom and discussing whether figures in the poppy field could be people if their eyes were not visible. Rather than positioning themselves as the holder of the 'right' answer and correcting the children straight away, the teacher encouraged the children to assess their own theories by inviting dialogue, as well as giving children opportunities to make their own observations about what happens when someone stands far away. The teacher also invited an artist who talked about perspective, for example how objects are smaller and may be painted more blurry to show they are far away.

The way in which children learn through Reggio Emilia education can also be likened to a broad practice called the negotiated curriculum (Beane, 1997), where teachers build learning opportunities onto children's interests and where the curriculum is co-constructed with children rather than prescribed. Negotiation affords children the possibility to experience autonomy because their voice and influence are valued (Seigel, 2010), and decisions are made in ways that they can be endorsed by individual children (Roche, 1996). Therefore learning activities and classroom decisions can be experienced as stemming from internal forces rather than imposed from the outside, and by following children's interests and curiosity, teachers can support children's intrinsic motivation in learning.

2.3 Autonomy as a Need and a Right

Whilst this thesis focuses on the experience of autonomy to support internally-motivated learning, it is also important to note that autonomy matters to education in other ways as well. Autonomy matters because it is part of children's socialisation into being autonomous beings. In other words, children benefit from being active and agentic learners (Begus & Bonawitz, 2021; Baker et al., 2021) and without opportunities to self-direct and have agency, they will not practice important self-regulation and executive function skills (Marulis et al., 2020). Indeed, their passive status may lead them to act automatically, without thought (Baker et al., 2021), reproduce 'the institution's voice' (Mick, 2011), and therefore not come into their own 'dispositional autonomy', that is the ability to lead a life according to one's wishes (Young, 1986).

However, children's autonomy is also an important consideration in its own right, without reference to potential benefits or future outcomes. In Self-Determination Theory, the experience of autonomy is related to wellbeing because autonomy is considered a fundamental human need whose satisfaction contributes to eudaimonic wellness or 'living well' (Ryan et al., 2008; Ryan, 2017), whereas its frustration leads to psychological ill-health (Ryan et al., 2006). Some have also argued that autonomy is constitutive of well-being in the sense that attitudinal pleasures (i.e. what an individual is pleased about, rather than sensory pleasure) contribute to a person's wellbeing if they are free, which is to say they come from autonomous decisions, aligned with what a person cares about (Cuypers & Haji, 2008).

Finally, there is a moral dimension to autonomy as a human right, in particular in its conceptualisation as self-rule, which entails freedom from domination or coercion by others (Ryan & Lynch, 2003). Since the adoption of the UN Convention on the Rights of the Child, children's right to be involved in decisions that concern them has been enshrined in international law (UNGA, 1989, article 12). In addition, children's vulnerability and reliance on adults means that the latter hold much power over children, and some have argued that this power is sometimes misused: humiliations and suffering are discounted because of children's subordination to adults (Langhout, 2005; Myles, 2015). For example, when students are not allowed to take off their blazer during heat waves because of strict rules around uniform, or when students are refused access to toilets during lesson time – both issues which recently created social media controversies in England – adults are using their power over children in ways that create discomfort and may be potentially distressing, in the name of maintaining order.

In other words, the moral standpoint taken here is that it cannot be good enough to talk about educational outcomes (such as test scores or improved learning) if it entails denying children their rights and ignoring their experience of educational practices. From this perspective, it becomes essential to understand classroom practice through the eyes of children. Autonomy is therefore central to such concerns because it relates to children's experience of their learning as something they are engaging in willingly rather than something they are pressured or coerced into, and allows us to question adults' use of power.

2.4 Conclusion to Chapter 2

In this chapter, I have drawn on Self-Determination Theory to show how fostering children's inner motivation to learn depends on attending to children's experience of autonomy in the classroom. Autonomy in this context relates to individuals experiencing situations as originating from within one's self rather than imposed by external forces and autonomy in the classroom means that students experience learning as agentic rather than alienating.

I then discussed how teachers may support autonomy in the classroom by drawing on research on autonomy-supportive teaching, which consists in strategies such as offering choices, providing rationales, and showing understanding for students' negative emotions. However, the place of choice in supporting autonomy is complex, as decontextualized choice does not increase motivation, and it may be that choice between options increases motivation by providing opportunities to pursue interests rather than because it supports feelings of autonomy. However, in the context of autonomy, choice must be understood as 'action choices', which is about self-direction and self-governance rather than 'picking' options. I have also briefly described child-centred pedagogies that explicitly support children's autonomy and inner motivation to learn in the early years of schooling – playful learning, Montessori and Reggio Emilia – and the kinds of practices teachers employ in those settings which foster children's autonomy and inner motivation. Finally, I have advocated for the importance of autonomy to children's lives and experience of schooling more generally, arguing that it relates to their wellbeing and to issues of power in the classroom.

Against this theoretical backdrop, in the first Part of this thesis I turn to our attempts to encourage teachers to increase children's autonomy in the classroom through the Stepping Stones professional development programme. In addition, in Part II, I examine an instrument

which emphasises children's experience of their learning and their inner motivation to engage in activities.

Part I - The Stepping Stones Programme - Supporting Teacher's Practices for Autonomy in the Early Primary Classroom

Chapter 3 - Methodology to Part I (RQ 1)

3.1 Overview of Chapter 3

This chapter provides the methodological framework as well as the methods used in Part I of this thesis, which was concerned with the work we carried out with teachers through the Stepping Stones professional development programme. Specifically, the aim of Part I is to understand what happened when teachers attempted to increase how much autonomy and ownership children had over their learning, focusing on teachers' practice. This methodology chapter presents the methodological background for the study, situating it within the broader landscape of methodological frameworks. Beginning by a brief overview of why we chose to focus on teacher inquiry as a form of professional development, I go on to discuss how the programme could be evaluated. Because the programme was initially developed within an experimental framework, I present both support and criticisms of RCTs. I then argue that in order to meet our goals of increased understanding, an interpretive realist approach resulting in ideographic knowledge is most appropriate. I then outline my research questions and describe the methods used in the study.

3.2 Methodological Background to Part I - From RCTs to Interpretive Research

3.2.1 Teacher Inquiry as Professional Development

One criticism of education research is that much of it does not relate to problems that matter to teachers nor suggests approaches that teachers find useful and flexible (Mitchell, 1999). As a result, education research has been accused of failing to tackle real classroom problems (Gore & Gitlin, 2004; Hargreaves, 1996; Hillage et al., 1998; Mitchell, 1999). Interventions have the potential to bridge this gap, but often they do not have real life impact outside of the intervention itself. One of the reasons for this is that large scale interventions are costly, and their rigidity means they are slow to change and cannot be easily tailored to the needs of participants (Schindler & Berry, 2017). In addition, one issue with many programmes is that by the time evaluations have been carried out, if negative results are found it is too late to change the programme and both time and resources have been wasted.

Supporting teacher inquiries tackles both of these issues because teachers pursue questions or test ideas relevant to their classrooms and adapt practice as needed. Therefore, this kind of approach involves both immediate relevance and flexibility. Studies suggest that professional development is more likely to lead to sustained change when teachers are treated as professionals, when they have opportunities to collaborate and when they have some amount of autonomy (e.g. Hargreaves et al., 2013) so that they are able to choose goals and adapt materials to meet their classrooms' needs, compared to 'top down' or 'transmission approaches' to teacher development (Hauge, 2019; Williamson & Morgan, 2009) which prescribe specific practices and lessons to which teachers have to strictly adhere. Teacher inquiry is understood here to be the process of questioning and experimenting with practices, reflecting and discussing on the effects of different practices, and considering what could be done next. It is through this process of inquiry that teachers can make research knowledge locally relevant (McIntyre, 2005).

Communities of Practice (Lave & Wenger, 1991) are particularly well suited to this aim. This well-established approach to professional development seeks to create collegial, supportive environments with a focus on reflection and peer learning with regards to practice (e.g. Palincsar et al., 1998; Perry et al., 1999; Vangrieken et al., 2017). Communities of Practice therefore offer an opportunity for teachers to test research knowledge in their own classrooms and transform it into usable and personalised theory (McIntyre, 2005).

In addition, there is also a need for congruence between the practices that are promoted through programmes, and the teachers' own learning experience (Swennen et al., 2008). Our research is based on the principle that children can be powerful agents in their own learning, and that they bring with them rich experiences and knowledge as well as personal desires, interests and needs (see Chapter 2). Teachers, too, come to research and professional development with a wealth of expertise, knowledge and skills as well as personal histories (Williamson & Morgan, 2009) – they are no more empty vessels to be filled than the children they teach. The need to treat teachers as professionals, to trust their expertise and to provide them with professional agency is at the moment particularly salient and pressing in our context, i.e. England (e.g. Rycroft-Smith & Dutaut, 2017). By engaging teachers in Communities of Practice, we wished to practice what we preach.

3.2.2 Understanding Change in Practice

RCTs and ‘What Works’ Approaches. An important concern for programme developers, users and policy makers is that programmes have the desired impact, and this requires evaluations of programmes. In a positivist framework, it is often argued that experimental methods such as RCTs are the best method of evaluation because they can establish causation (in the sense that the programme caused a particular set of outcomes) and avoid many threats to validity (e.g. Trochim, 2016). In particular, RCTs rely on the random allocation of individuals to different groups receiving different programmes or interventions (often called ‘treatment’ groups or control groups because of the origin of RCTs in clinical research). This is in contrast to many studies investigating correlations between variables, which cannot show clear causal links between those variables (Wyse & Torgerson, 2017). This has led some to criticise the value of education research for practice and policy (e.g. Hargreaves, 1996) and to promote RCTs as the ‘gold standard’ in evaluations (EIF, 2015; Goldacre, 2013; Mertens, 2015). However, RCTs are often difficult in education; for example, children from the same classroom cannot be randomly assigned a ‘treatment’ group if the intervention relies on the teacher’s practice. As a result, variations of RCTs (called Quasi-Experimental Designs or QEDs) are sometimes used in which there is a control group but not random assignment (Trochim, 2016). However, for brevity I refer to the RCTs for the remainder of this chapter to also include studies that use QEDs.

Because RCTs are costly, programmes are not initially ready to undergo such testing when they are first being developed, and instead must first go through a process of increasingly robust forms of evaluation. For example, the EIF, which is responsible in England for evaluating what Early Years interventions ‘work’ and advise government, has described a multi-stage framework for evaluating early childhood programmes (EIF, 2015). At early stages, programmes develop robust theories of change and the framework culminates in multiple, independently-evaluated RCTs. At the time of implementation, the Stepping Stones project aimed to follow such a framework and build up towards a large scale RCT evaluation. At early stages of development, however, evaluation initially focuses on feasibility rather than impact. In other words, programmes must be found to be deliverable before they can be found to be effective, and feasibility studies involve asking, ‘Can this intervention be done?’ (Hoddinott, 2015). Often feasibility concerns the content (‘what’) and delivery (‘how’) of the intervention; this is akin to process evaluation.

Process evaluation, as opposed to impact evaluation, concerns questions of how interventions operate, for example in different contexts, rather than whether they are effective, and typically rely on qualitative methods of data collection (Mertens, 2015). Although process evaluation can be carried out at any stage of intervention development and testing, including to add validity to RCTs (Green & Glasgow, 2006; Siddiqui et al., 2018), it plays a particularly important role in feasibility studies. Nonetheless, in this context both feasibility studies and process evaluation are seen as preliminary steps to the ‘final product’ of a programme, which is then tested through an RCT.

Criticisms of RCTs in a Complex World. Whilst RCTs are still seen by many researchers, policy makers and evaluation bodies (e.g. EIF) as the ‘gold standard’ of evidence on which education decisions should be based, an increasing number of education researchers have begun to doubt and criticise this position (Biesta, 2007; Burnett, 2017; Simpson, 2017; Wrigley, 2018). Whilst the general project was framed as paving the way towards an RCT, there were many aspects of the programme which did not fit well with the requirements and framework of RCTs. In particular, our use of teacher inquiry meant that what teachers implemented in their classrooms varied widely. In addition, the nature of what we were interested in promoting – children’s autonomy – is itself complex, nuanced and as we will see contextual. I briefly outline why this was incongruent with an RCT framework.

First, in order to show strong effects, RCTs need to be uniform and therefore rigid – what Burnett (2017) calls “fixed, linear logic” (p. 2) – and therefore are not well-suited to practices that adapt to the needs of children, and result from moment-by-moment decisions made by teachers. Even if teachers follow a rigid script (and therefore all children receive the same practice), children may not all interpret it in the same way; in other words, the ‘treatment’ may not only not have the same effect on everyone, it might not even be the same ‘treatment’ in the first place. Similarly, what constitutes the ‘control’ group may create issues – the control group may receive ‘business-as-usual’ practices that are not so different from the ‘treatment’, and there may be pre-existing differences between groups. Whilst RCTs aim to eliminate such systematic differences between groups, this is difficult in practice (see Wrigley, 2018 for a more detailed example).

Secondly, as a result of the first issue, programmes are often applied unevenly. For example, a recent report on teacher professional development RCTs found that “even faithfully-implemented programmes see substantial adaptation and deviation from plans” (Sims et al., 2021, p. 45). Teachers may adapt programmes to respond to children’s needs or

local constraints, and their practice may vary from one moment to the next, therefore making it difficult to show fidelity and impact. In addition, much of teaching is complex and nuanced, and therefore it is difficult for a programme to be a single object (e.g. Ellaway et al., 2014) – in other words, any programme in education is threatened by issues of construct validity (Trochim, 2016). Seen in a more positive light, education interventions need to be understood in terms of multiplicities, in that they are constituted and interpreted differently each time they are applied (Burnett, 2017). This is not a matter of solving the problem by making delivery more rigid and more prescribed, but rather that the conceptualisation of a programme as a single object – what Law (2004) calls singularity – denies the complexity of real classrooms and the associated responsiveness and flexibility teachers must therefore demonstrate (Ellis & Moss, 2014).

Thirdly, RCTs are not sensitive to variations in the programme's impact, because they only provide an average effect (Pawson & Tilley, 1997). Therefore, a programme could be highly effective for some individuals, and highly ineffective (or damaging) for others, but only an average effect would be reported. This links to the criticism that RCTs fail to improve our understanding of how programmes work or why the problem exists in the first place (Pawson & Tilley, 1997; Wrigley, 2018). As a result, impact tends to be small and many RCTs and meta-analyses of programmes fail to show any impact at all (Bowers, 2020; Lortie-Forgues & Inglis, 2019; Pawson & Tilley, 1997). Whilst the move away from 'what works' to 'what works, for whom, under what circumstances' may answer some of these issues, this may take a variety of forms which may sit more or less well with experimental designs (e.g. segmentation, Chong et al., 2019; realist evaluations, Pawson & Tilley, 1997).

These issues relate to broader criticisms of attempts to prove 'what works' in education. In particular, critics argue that the 'what works' approaches and the 'evidence-based' discourse displace efforts, funding and discussions away from considering the 'good' of education more broadly and what education is for (Biesta, 2010) as well as from more complex and nuanced programmes that may not easily show statistical effects (Burnett & Coldwell, 2020). In addition, its technical and managerial focus and top down nature (teachers are told what good evidence is and what techniques to apply) are in tension with the recognition described above that teachers are knowledge producers and need to adapt evidence to their own contexts (McIntyre, 2005), as well as the conception of teaching as requiring both technical and craft knowledge (*ibid*) or phronesis (Biesta, 2010). These approaches therefore make reductive assumptions about what should count as evidence about education as well as about the nature of professional development (see Biesta, 2010 for

further discussion). This is not to say that RCTs do not have a role to play in education under certain circumstances nor that some of the issues presented here cannot be mitigated in well-designed studies (see Wyse and Torgerson, 2017). However, the general framework under which RCTs operate is at odds with the present research, and in particular with the collaborative inquiry-based approach to professional development we took as well as with the focus on children's autonomy. Even early stages that would not rely on experimental methods are nonetheless part of the same way of conceptualising impact and evaluation. For example, feasibility studies focus on superficial though practical aspects of the programme, such as whether it is desirable or acceptable to users, rather than on building understanding of how the programme is interpreted and enacted.

3.2.3 Interpretive Realist Research

Biesta (2010) argues that research does not only play a role in finding technical solutions to problems teachers may encounter in the classroom, but also a cultural role in providing lenses through which teachers may view their practice. Such a view of the role of research fits well with a view of professional development as not merely technical but as solving unique problems in action (Biesta, 2010; McIntyre, 2005). Biesta argues that such research has practical implications because it can help us “understand problems we may not have understood before” and “to envisage opportunities for action where we did not envisage them before” (Biesta, 2010, p. 45). In examining the experiences and actions of teachers in the Stepping Stones programme, it is such understanding and envisaging that I am seeking, rather than proving the effectiveness or feasibility of the programme. I believe it is the former, rather than the latter, which will be most useful for our understanding of how to support teachers' practice towards greater autonomy for children.

In order to understand teachers' change in practice as a result of the Stepping Stones programme, and more generally how teachers can support children's autonomy in the classroom, I therefore take an interpretive realist approach. It is interpretive because it relies on interpretations (by the teachers and the researchers) to create understanding. Whilst all research, even quantitative and positivist research, involves interpretation, interpretive research explicitly acknowledges the contribution of interpretation in analyses. However, I do not mean this research is interpretive in an ontological sense, which would imply social reality is constructed, and does not exist outside of these constructions. Instead it is interpretive in a methodological sense. To understand this distinction, I draw on realists such

as Maxwell (2012a), who have argued that it is important to separate ontology and epistemology, and therefore it is possible to have ontological realism but epistemological constructivism. When the two are conflated – a term realists call the epistemic fallacy – then all of social reality is constructed. The difficulty with such a position is that if there is no reality or truth to be found (or approached), social realities are “willed into existence” (Burrell & Morgan, 1979, p. 233) rather than pre-existing, and so there is little scope for arguing for one interpretation over others. Such relativism makes all knowledge a form of opinion, all on equal footing (Blackburn, 2005), in which no-one can claim to have a better explanation or representation of reality. Where realists differ from (ontological) interpretivists is in their argument that social reality, whilst it may be bound up with human constructions and enacted through human agents, also has a reality outside of those agents. In particular, Archer (1995) makes a strong case for why social structures have causal influences – and therefore are real – and cannot be reduced to the agents they apply to. Nonetheless, this ontological realism accepts that social reality may be transient and is not governed by universal laws because it is part of open systems (Sayer, 1997). In addition, our knowledge of it is always partial and constructed, and therefore socially bound (epistemic constructivism) (Maxwell, 2012a).

Therefore, the separation of ontology and epistemology means that a realist ontology is not incompatible with qualitative research that fully embraces the constructivist nature of such research – what Braun & Clarke (2021a, 2021b) call Big Q research. Indeed, these qualitative researchers acknowledge the potential of (critical) realism to underpin interpretive research. In this research, because we were concerned with understanding, at a deep level, what happened when teachers introduced greater autonomy for children, an interpretive methodology which allowed sense-making (and not just hypothesis-testing) was important. The knowledge produced as a result of such work, as Yanow (2015) argues, is “understood as being produced not through disembodied reason but through the situated context of the ‘knower’ producing it” (p. 10). Where this research perhaps diverges from ‘classical’ interpretive research (e.g. Clarke and colleagues) is that it is interpretive, but not phenomenological. It is interpretive in that it is concerned with building understanding and acknowledges the role of the researcher in creating that understanding; therefore the theories do not make any claim to certainty. They are arguments, built to explain the data. But it is not phenomenological because it is not a representation of the teachers’ subjective experiences with taking part in the programme (i.e. ‘what it was like for them’). Although this forms part of the analyses, I also describe and explain what happened in the classrooms, through the lens

of children's autonomy in learning ('what the teachers did') and how this related to norms, discourses and structures in the classroom – therefore, attempting to provide some explanations as to *why* events occurred as they did. Teachers' experiences form part of those explanations because their experiences are related to what happened, but the latter is not reducible to teachers' subjectivity.

3.3 Part I Research Questions (R.Q. 1)

The overarching aim of this part of the thesis was to understand what happened when teachers attempted to increase children's autonomy in the classroom, understood here to encompass not only choice(s) but also ownership over the learning and opportunities to make and participate in decisions about the learning (as per Chapter 2; see also Reeve et al., 2003). Specifically, I was interested in investigating the following question:

R.Q. 1 What changes to their practice can teachers implement by taking part in a professional development programme focusing on increasing children's autonomy in the classroom?

There is no one-to-one correspondence between programme activities and what teachers do in their classrooms. Instead, teachers need to make sense of the materials and adapt it to their own classrooms. The kinds of changes they made, the extent to which they were able to change their practice, and what affected their ability to do so, were the three areas of focus. However, it is important to note these three sub-questions were not investigated independently of each other; instead, they formed different facets of the same question of understanding teachers' change in practice with regards to children's autonomy in the classroom. In addition, by investigating teachers' change in practice towards greater autonomy for children, we sought to investigate how teachers' support of children's autonomy is enacted in real classrooms, and therefore contribute to discussions around teachers' autonomy-supportive practice (as opposed to teacher professional development per se).

R.Q. 1.1 What opportunities for autonomy did teachers provide for the students in their classrooms?

In order to understand what changes teachers made to their practice, we need to understand what their practice was like before, as well as throughout, the programme. We anticipated that teachers would have different starting points, and we expected that most teachers would already be providing some forms of autonomy support. In addition, teachers needed to translate and integrate the content of the development programme into their own practice. We therefore predicted teachers would interpret the materials in different ways and that context would play an important role in how they enacted the ideas we had presented in the workshops. The aim of this question was to focus our attention on the kinds of changes to the classroom life that teachers made as well as the ways in which they supported or constrained children's autonomy.

R.Q. 1.2 To what extent did teachers change their practice through the intervention programme?

The focus of this question was on how much of a change teachers were able to make, rather than the form it has taken. Although this seems to have a quantitative slant, this question was inter-related with the question of how teachers translated the development programme into real practice and therefore purely investigated qualitatively: teachers may be able to make large changes for some aspects of their practice, and only limited changes in different areas. Therefore, both RQ 1.1 and 1.2 were understood as investigating the nature of the change in teachers' practice. Specifically, RQ 1.2 included considerations around how prevalent and embedded the new practices had become in relation to the rest of the teacher's practice; for example, whether the new practices were restricted to some areas of the curriculum or had spilled over to other areas, as well as how transformative the changes had been to the teacher's overall practice.

R.Q. 1.3 What factors, including resources, structures and aspects of the development programme, have affected the ease or difficulty of teachers modifying their practice?

This question was concerned with aspects of the programme but also the broader structures around teachers that hinder or facilitate the process of changing teachers' practice and increasing children's autonomy. This question sought to address more than just teachers' satisfaction about different aspects of the development programme, as might be found in a

satisfaction survey or a feasibility study, and instead to understand the relationships between the different components of this system (the programme, the teachers and their context). For example, we anticipated there would be tensions between the prescribed curriculum and giving children more say in choosing the direction learning takes. The aim of this question was to go beyond a list of constraints, and instead to explore how teachers navigated those difficulties, what conditions gave rise to them, and what supported teachers in overcoming them, both within the programme and within the school context. We also anticipated we might uncover situations in which it would not be possible for teachers to change their practice because of such constraints and difficulties and sought to examine what made them so.

Whilst the research questions remained largely unchanged over the course of the project and subsequent analyses, their meaning subtly shifted from one of evaluation to one of understanding, as the goal changed from evaluating the feasibility of the programme to creating understanding of teachers' practice. This is an important distinction as this shift occurred alongside a refinement of the methodology, away from positivist accounts of evaluation to embracing a fully interpretivist account of change. It also shifted the focus away from whether or not the programme was desirable or had potential (including what barriers existed) to understanding dynamics in the classroom that interacted with the goals of the programme – children's autonomy – and the ways in which teachers enacted and grappled with those in the classroom. Finally, it moved away from holding the programme to be a singular object (to be evaluated) to considering its multiplicity in different contexts: the emphasis thus shifted away from the programme itself and the strategies teachers used, and instead focused towards the many practices of how teachers might encourage and open up spaces for children's autonomy in the early Primary classroom.

3.4 Methods for Part I

3.4.1 Overview of the 2019 Stepping Stones Professional Development Programme

The aim of the Stepping Stones programme was to help teachers increase the amount of autonomy they provided to children in their classrooms by inviting them to experiment with their practice. Initial aspects had been previously piloted with a group of eight teachers in 2016-2017 (see Kittredge et al., 2018), and the input of various partner teachers had been sought at other points in the development of the programme. Whilst the development of the

programme has been continuous, the work carried out with teachers described in this thesis took place between March and June 2019. Hereafter, I refer to this work carried out with teachers in 2019 as the Stepping Stones *programme*, and continue referring to the Stepping Stones *project* as the larger research project around it, including, for example, the theoretical work described in Chapter 2 and the work described in Part II.

The Stepping Stones programme took the form of a Community of Practice organised by the research team where teachers met four times throughout the programme to reflect on their practice, discuss and exchange ideas, and receive input from the research team. Although the overarching aim of the programme was set by the researchers, from the onset the teachers were invited to be active contributors to the programme, with the aim of empowering them in the changes they were making. We introduced their role as that of co-researchers, which we described as taking a stance of inquiry towards their practice. The iteration of workshops meant that teachers were able to set their own goals, tinker with their practice and reflect on these changes.

3.4.2 Participating Teachers

Nine teachers participated in the Stepping Stones programme (Table 1). Two teachers also initially participated in the programme but dropped out from data collection for reasons external to the study (maternity leave and bereavement) before the final workshop, and their data were therefore not included in the analyses. One additional teacher joined the programme late and attended the initial workshop, but a change in her school and her teaching duties meant she was not able to attend future meetings. No data were collected from this teacher. The analyses therefore concern the nine teachers who were involved throughout the programme and from whom data was collected. They came from five Primary schools in Cambridgeshire, Suffolk and Norfolk (Schools 1A-1E).

Classroom Characteristics. All teachers taught Year 1 children (aged 5-6 years) at some point in their week, but a number of teachers also taught other classes, and some classes were mixed-age groups. Some teachers in the programme taught the same classes part time. Only three teachers taught Year 1-only classes full time. Most teachers came from one-form and two-form entry schools, and two teachers came from a small rural school with only two classrooms from Reception to Year 6.

Table 1 *Classroom Characteristics by Teacher in the Stepping Stones Programme*

Pseudonym	Years of experience	Number of classes per year group in the school	Mixed-age or single-age classes	Part time or full time teaching	Percentage bands ^b of children eligible for deprivation premium (schoolwide, 2019)
Danielle	>10	1	Single-age	Full time	20-30
Vicky	>10	<1	Mixed-age	Part time	10-20
Anne	>10	<1	Mixed-age	Part time	10-20
Lisa	>10	2	Single-age	Full time	10-20
Beth	9	2	Single-age	Full time	<10
Julie	4 + 11 years as a TA ^a	2	Single-age	Full time	10-20
Claire	3	2	Single-age	Full time	10-20
Stephanie	>10	<1	Mixed-age	Part time	20-30
Helen	N/A	<1	Mixed-age	Part time	20-30

^a A TA is a Teaching Assistant, a common supporting role in English classrooms, requiring a lower level of qualification than teachers; ^b The percentage of children eligible for “Pupil Premium” is generally considered in England to be an indicator of socio-economic deprivation, with a national average in 2019 of 15%. It is represented here in the form of bands to preserve schools’ anonymity.

Recruitment. Teachers were recruited through the team’s existing networks and previous contacts. Four of the nine initial teachers were recruited through existing relationships with their headteacher. One teacher was recruited through word of mouth from one of these headteachers, and the remaining four teachers were recruited through an advertisement to mentors in the University’s teacher training programme.

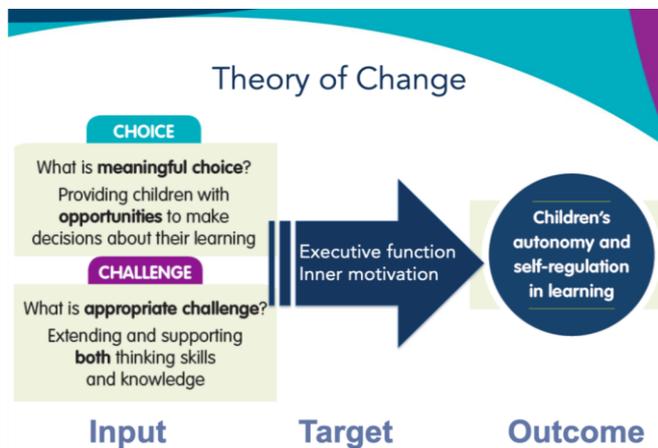
After initial recruitment, contact was made through the teachers directly rather than through the headteachers. Teachers were given information about the programme and opportunities to refuse our invitation to take part in the programme – indeed one teacher who was initially invited decided not to participate, in addition to those teachers who dropped out because of unforeseen circumstances. Teachers were also given a Memorandum of Understanding describing the programme and its collaborative nature, what it would require of the participants, as well as our duties as researchers and the safeguards we would put in place (Appendix A). Therefore, despite many of them having been recruited through their managers, we feel teachers were not pressured into taking part, and this was also checked again in the interview through a question on how they had become involved in the programme and how they felt about it.

3.4.3 Content of the Programme

Programme Aims & Theory of Change. The aim of the programme was informed by research on the role of feelings of autonomy for motivating learning, as described in Chapter 2, as well as the importance of providing children with opportunities to challenge their executive functions (e.g. Diamond & Ling, 2016), based on work carried out by the Principal Investigator and other researchers in the team. Based on this work, we developed a Theory of Change (Figure 4) identifying the kinds of teaching practices that could lead to our desired outcomes. In particular we focused on the notions of *choice* and *challenge* as important ingredients to support children’s executive function and inner motivation. I have extensively reviewed the importance of choice for children’s autonomy and inner motivation in Chapter 2. The role of challenge came from literature on the importance of optimal challenge for inner motivation in flow theory (e.g. Abuhamdeh & Csikszentmihalyi, 2012a), the role of perceived difficulty in motivation (e.g. Fulmer & Tulis, 2013; Patall et al., 2018), the importance of challenging executive functions in order to improve them (Diamond & Ling, 2016), as well as broader learning theories, including the Zone of Proximal Development originally attributed to Vygotsky (Hedges, 2021). However, as I discuss later, challenge is less clearly linked to the experience of autonomy and therefore was not central to the analyses presented in this thesis in the same way as choice. Nonetheless, it is important to note that by *challenge* we did not simply mean teachers increasing the level of difficulty of an activity for children who had completed a task or were able to do more difficult work. Instead, providing an optimal amount of challenge was as much about supportive systems

and interactions as it was about extending children’s thinking and providing opportunities for students to challenge their mental capacities through self-direction.

Figure 4 *A Theory of Change for Developing Children’s Autonomy in the Classroom*



To make choice and challenge more concrete, we presented the teachers with 11 practices or ‘strategies’ (Figure 5 – see also Appendix B for how we came to these strategies and Appendix C for full details).

Figure 5 *Strategies for Choice and Challenge Shared with Teachers*

- Show flexibility in planning to follow children’s interests**
- Elaborate on children’s answers**
- Ask open-ended questions**
- Set up activities as continuous provision (“free choice”)**
- Involve children in deciding learning goals**
- Help children evaluate successful strategies and each other’s work**
- Provide open-ended tasks**
- Provide strategies for independent problem-solving**
- Allow differing opinions**
- Model and encourage empathy**
- Observe, listen & notice engagement**

The first three strategies, initially presented separately as introductory strategies, were illustrated by written examples of real teachers’ practice derived from research and

pedagogical texts (Appendix C1). The aim of these texts was not to provide perfect examples to copy, but rather to engage teachers in reflection about how the principles we presented could be enacted, whilst engaging with the complexities of real situations (Korsgaard, 2020). Drawing on these practices, teachers were then invited to set themselves goals of what they would like to try in their own classroom, depending on their context, needs they had identified and their own vision for their classroom.

The Community of Practice. The teachers took part in four Community of Practice workshops over ten weeks (see Appendix D for detailed descriptions of each workshop). The first workshop was a day-long introductory workshop and each subsequent workshop lasted around three hours. Workshops took place at the Faculty of Education, University of Cambridge.

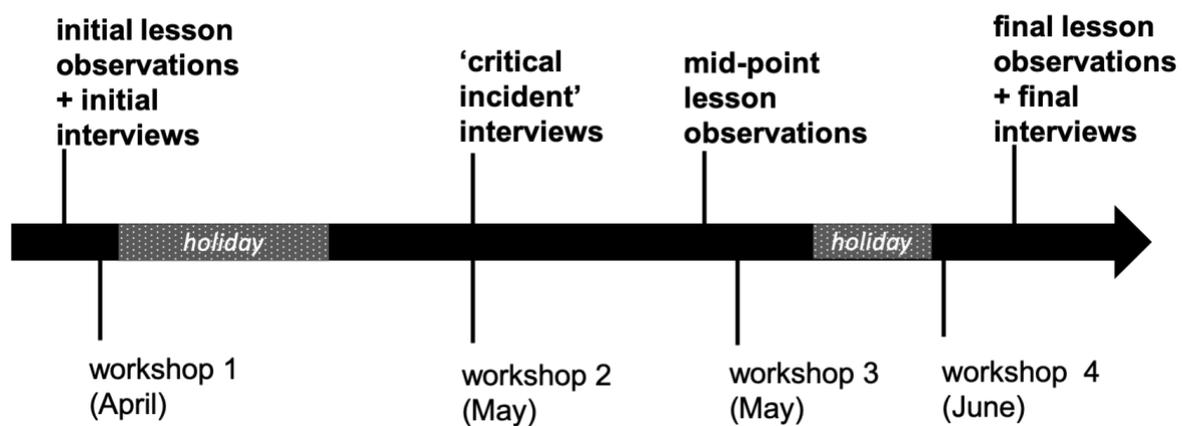
The first workshop consisted of an introduction to our Theory of Change and discussions of examples of teaching practices, as well as a planning period. This occurred the week before the Easter holidays, to allow teachers to take stock of the ideas they had been introduced to, and to plan for the following term. In addition, in order to find a suitable date, the teachers attended the initial workshop in two groups of five and seven (respectively four and five of the final participants who completed the programme).

The next three workshops involved time for individual and shared reflections, additional support from the research team and time for planning and goal setting. We aimed to strike a balance between providing the teachers with enough support, models and theoretical understanding so that they felt they had the tools to take action, whilst having the autonomy to take the research in the direction that suited their interests and their context. Overall, the Community of Practice was organised around a ‘Plan, Do, Review’ model (e.g. Baumfield, 2008) where teachers were encouraged to plan how they might implement a strategy, experiment with it in their classroom, then review and evaluate what had occurred as a result in discussions and activities during the workshops (see Appendix D for further details on activities). To help with reflections, teachers were encouraged to bring plans and evidence into the workshops, and teachers sometimes brought pictures of the children’s work, though not all teachers did. However, whilst reflections were supported during workshops using question prompts and templates, teachers’ inquiries in their classrooms were unguided. In addition, three of the nine teachers were unable to attend the final workshop, and no alternative date could be found – these teachers received the material we gave out but did not benefit from the discussions.

3.4.4 Data Collection

Data for this part of the thesis came from three sources: interviews, lesson observations and Community of Practice workshop discussions at different time points (Figure 6).

Figure 6 *Timeline of the 2019 Stepping Stones Programme*



Note. The programme took place over 10 weeks from the first to the last Community of Practice workshop. Data collection took place on either side of this period as well as during the programme. Holidays refer to school holidays during which teachers were not in school.

Interviews. We conducted semi-structured interviews at the start and end of the programme to investigate the development of the teachers' understanding and practice, as well as short five-minute "critical incident" (CI) interviews during the second Community of Practice workshop. Initial interviews took place in the week before the first workshop. Most final interviews took place on the same day or shortly after the final class visit. However, one teacher was unable to do her post-programme interview on the same day as the class visit, and then it had to be postponed because of medical reasons; the interview thus took place six weeks after the final workshop. In total, we carried out 18 long interviews (which lasted between 15 minutes and an hour) and 9 five-minute 'CI' interviews.

The initial interview focused on teachers' backgrounds, their existing practice, and in particular their thoughts on choice and challenge (Appendix E1). The five-minute interviews focused on 'critical incidents' (Appendix E2) as per Kain (2003). In this type of interview, participants are asked to recount an event that is significant in some way. In our programme, we asked teachers to think of a time when they had tried out one of the strategies and which

stood out to them in terms of how easy or difficult it had been. The final interview focused on the strategies the teachers had tried out, their thoughts on choice and challenge, and their general experience of taking part in the programme (Appendix E3). The interview also included two structured tasks using a ‘think out loud’ method (Appendices E4 and E5). One task asked teachers to rank the different strategies depending on their level of difficulty using a Q-sort frame (Brown, 1993). The second task asked teachers to rank barriers to giving children greater autonomy; the teachers were provided with the barriers that emerged at the third Community of Practice workshop, but could discard barriers irrelevant to their context and add new ones as they saw fit. The tasks were not formally analysed but the rankings are discussed in the analyses and the teachers’ thought processes were analysed as part of the interviews.

Lesson observations. We conducted naturalistic classroom observations at three time points throughout the programme: before the first workshop, in the middle of the programme and after the last workshop. In order to record these observations, we used running records (Appendix F), which are in-the-moment, detailed descriptions of events as they unfold (Perry & VandeKamp, 2000). The aim of these observations was to observe teachers’ actual practice, rather than only the practice they reported, to base analyses not only on teacher’s interpretations but also on actual events, in order to achieve descriptive validity (Maxwell, 2012a). However, whilst running records should be as neutral in tone as possible, it is impossible to record everything that happens in a lesson, and the researcher’s attention is inevitably guided by the lens through which she views classroom practice. Therefore, to ensure we also captured the teacher’s perspective on the lesson, the observations were followed up by a short debrief focused on the teacher’s intentions for the lesson and what they had noticed. It was also an opportunity to clarify specific events. As well as providing an important source of data, these debriefs aimed to restore the balance in our relationship with the teachers. Being observed by a third party can lead to feelings of disempowerment, and this was a chance for teachers to regain power by telling their own story about the lesson. At each observation point (Figure 6), two lessons were observed for each teacher, resulting in a total of 54 running records. Teachers were only observed in classes where they taught at least some Year 1 children. An effort was made to observe one ‘core’ lesson (literacy or numeracy) and one other lesson, but this was not always possible because of timetables, school events, and the limited timeframe within which to observe all teachers at each time

point. For 11 out of the 54 lessons, it was not possible to do a debrief with the teachers after the lesson, so only 43 debriefs were recorded.

Workshop Discussions. The workshop discussions were not routinely recorded because of difficulties in previous projects the research team had undertaken: effort in transcribing often long conversations, difficulties in identifying the speaker and participants speaking over each other. However, two discussions were exceptionally recorded because of their relevance:

- a discussion in the 3rd Community of Practice meeting (Workshop 3) around an activity where teachers recorded challenges to implementing the strategies;
- and a discussion in the 4th Community of Practice meeting (Workshop 4) where teachers discussed some changes they had made to their practice, and where one teacher was invited to share a relevant aspect of her practice, continuous provision.

Transcription. All debriefs and interviews were audio-recorded. All audio files were transcribed into a written format. This was done by three different researchers. Running records, which were already electronic word documents, were edited for typos and any abbreviations filled in.

Training and protocols. Data were collected by two researchers: myself and a Research Assistant; a third researcher also helped carry out CI interviews. We followed the same teachers throughout the programme both for practical reasons and to ensure continuity (five and four teachers respectively). Interview schedules and guidance as well as running record templates were created in advance to maximise consistency as far as possible but also to leave room to follow conversations where they led. We practiced the interview questions and running records together in a Reception classroom and with a Year 1 teacher in Cambridge, and later reviewed and edited the documents together to help with consistency.

3.4.5 Coding and Analysis

I conducted qualitative analyses on all three sources of data using an inductive approach, drawing on reflexive thematic analysis (Braun & Clarke, 2006, 2019) as well as using a more narrative approach to each teacher's experience of the programme using stories of change.

Contiguity-Based Narrative Techniques. There is increasing recognition that change in practice is not uniform or linear, and that research needs to be able to show these trajectories in order to truly capture the effect of a programme (e.g. Peressini et al., 2004; Vermunt et al., 2019). Treating change as something individual, that differs from person to person, is also part of the recognition that there are important lessons to be learned not only in the overall patterns within a group but also in the uniqueness of each story. Maxwell (2012a) discusses the benefits and limitations of what he calls *categorising* and *connecting* strategies. Coding, including through thematic analysis, splits up the data, and in so doing loses the contextual relationships that are essential to understanding context-dependent mechanisms. Maxwell distinguishes *similarity* from *contiguity*. Categorising, he argues, allows us to find *similarities* across situations, but does not allow us to understand the sequence of events through time and space (contiguity). Narrative techniques, on the other hand, are based on contiguity. Therefore, in order to first capture teachers' idiosyncratic changes to their practice and maintain their relationship to context, teachers' change in practice was first analysed as 'stories of change'. The stories of change were written by focusing on one teacher at a time, and recounting their attempts to try new strategies using all sources of data. Here I present four teachers' stories of change, chosen because they represented a wide range of contexts and changes. Two teachers were chosen because they made larger changes to their practice: one was using continuous provision, which required an organisational change, whilst another followed children's interests in her planning. Two teachers were chosen because they had made much smaller changes: one struggled with the programme, whilst the other was required to follow very strict lesson plans and therefore had little freedom to make changes to her practice.

Despite focusing on contiguity, narratives can also include functional categories – like parts of a story, these categories structure the narrative into functional units whilst maintaining the sequence of events (Maxwell, 2012a). This is the case here, as the stories of change are split into similar sections (teachers' initial practice, teachers' attempts) and then further split into the *kinds* of practices and the *kinds* of changes each teacher made. There is therefore some similarity across the stories, though what the teachers tried and their responses to it varied, and the story is told in its rich context.

However, using narratives alone makes it difficult to see what was common across the entire programme, and to uncover underlying mechanisms. In other words, it can make it difficult to gain insights from different contexts and to see the bigger picture (Maxwell,

2012a), and this made a categorising analysis necessary. In addition, the stories of change were not written independently of the thematic analysis – indeed, the emerging themes from the coding (described below) guided the focus of the stories of change. In return, the stories of change allowed me to stay grounded in the context of each classroom. As Maxwell argues, categorising and connecting strategies should be seen as complementary and mutually supporting (Maxwell, 2012a).

Reflexive Thematic Analysis. For the categorising phase of the analysis, I drew on thematic analysis, though had to adapt it because of the nature of the research, which investigated change and mechanisms from a realist interpretive perspective rather than participants' subjective experiences. Thematic analysis can take many forms (Braun & Clarke, 2019), but here I focused on the one described by Braun and Clarke (2006, 2019, 2021a, 2021b) which they have characterised as reflexive thematic analysis. It is an inductive and interpretive form of thematic analysis, where themes are “stories about particular patterns of shared meaning across the dataset” (Braun & Clarke, 2019, p. 592) and where the categories are understood to be actively constructed by the researchers rather than emerging from the data (Braun & Clarke, 2019). Whilst Braun and Clarke emphasise the interpretivist (Big Q) nature of reflexive thematic analysis, they also acknowledge that it can be compatible with realist ontologies (Braun & Clarke, 2021a, 2021b). It was therefore relevant here, as it allowed us to understand teachers' changes in practice, and in particular what affected their ability to make changes. In particular, using a latent approach – going beyond the surface of what teachers said or did to identify underlying ideas and conceptualisations (Braun & Clarke, 2006) – allowed me to theorise why certain teachers had responded to the programme in certain ways. In this regard, however, it may be different from Braun and Clarke's conceptualisations of themes, which is concerned with patterns of shared meaning, rather than mechanisms. Thematic analysis has been predominantly used to understand participants' experiences or ideas around certain issues, rather than their actions – in other words, it traditionally focuses the explanations participants give, rather than causal mechanisms inferred through observation and abduction.

Therefore, the method of reflexive thematic analysis was slightly adapted for the present study but retained its important features. It involves six steps (Braun & Clarke, 2006; 2019): familiarisation with data, open coding, searching for themes, reviewing themes, refining and naming themes, and writing up (see Appendix G for more details). Initial coding was carried out in NVivo (QSR International, 1999). I deviated from Braun and Clarke's

method in the later stages of analyses. In particular, I used a visual map for searching for themes using an online software (Kumu Relationship Mapping Software, 2020), where I added all codes generated in the initial stage and established relationships between them. This stage of coding resembled more what is known as ‘axial coding’ in Grounded Theory (Glaser & Strauss, 1967), where analytical linkages are made between codes of data (Birks & Mills, 2015). Grounded Theory as a general methodology was not appropriate, in particular its reliance on theoretical sampling, as the programme was a time-limited programme which did not lend itself to iterations of sampling and simultaneous analysis. However, some of the methods of Grounded Theory are helpful for making sense of collected data (which Braun and Clarke, 2006, call ‘Grounded Theory-lite’, p. 81).

In addition, early on the themes generated were more akin to what Braun and Clarke (2019) call domains summaries, for example around types of barriers and benefits. This more superficial level of analysis was later on replaced by the final ‘story’ presented here. In addition, whilst coding was inductive, it was not atheoretical: I did not impose categories on the data, but the researcher always brings their worldview to their analysis (Braun & Clarke, 2019). Theories, including our Theory of Change, as well as assumptions (in particular with regards to the purposes of education and the importance of autonomy – see Chapters 1 and 2) coloured the interpretations I made of the data. Nonetheless, coding was carried out without mapping those theories onto the data. For example, I did not code the data for the use of strategies we had introduced, as this would have narrowed the outcome of the analyses artificially – in other words, it would only have answered the question of which strategies the teachers used, but would not have shed light on the other dynamics within the classroom. In addition, it was not necessarily the case that the strategies we had presented were the best way to understand teachers’ practice. However, links to the theory and wider research were still made, in particular at later stages of coding and during writing up. In particular, I relied on abduction as a logical tool. Abduction involves inference to the best explanation (Maxwell, 2012a; Shurz, 2008), and therefore examining the strength of evidence for different theories (see also Appendix H on validity in interpretive research).

3.4.6 Further Ethical Considerations

This research was subject to the Faculty of Education ethics procedure and followed BERA guidelines (British Education Research Association, 2011). Here I discuss

considerations regarding potential detriments, informed consent, and confidentiality, as well as ethical questions around the position of the teachers in this research.

Potential Detriments. Potential detriments may occur for teachers through demands on their time. Teachers are very busy, and it is therefore especially important that the programme is a good use of their time. Communities of Practice are particularly suited to the needs of teachers because they help teachers put into practice what they are learning; they also help teachers become more reflective and they can apply this skill to other areas of their practice. However, it was important that the time teachers spend on the programme did not encroach on their personal lives or their other duties. Therefore an important aspect of the Community of Practice workshops was that they happened during school time. Funding through the Stepping Stones programme paid for cover to release teachers from their classroom at no financial cost to the schools.

There could also have been detriments to the children if the programme negatively impacted on their learning. To minimise this risk, the intervention was based on a robust Theory of Change and teachers were in a position where they had control over what they implemented, so the children's needs could be prioritised over the programme if it was found to impact children negatively.

Consent. All children received their teachers' changes in practice. Teachers reflecting on their practice, attending training and changing elements of their practice is a natural part of their work, and schools are considered to have the right over such decisions. However, to ensure parents understood the research and felt comfortable with the programme, it was important to ensure good communication with them. We followed the schools' lead on what was most appropriate for their community. In many cases, an information letter was sent home with the children.

As previously discussed, teachers and researchers both signed a Memorandum of Understanding (Appendix A). This explicitly asked for teachers' consent to the research and informed them of their right to withdraw from the study at any time.

Privacy and Confidentiality. BERA guidelines (British Education Research Association, 2011) advise on the need to respect both participants' privacy and their right to authorship. This is difficult in qualitative research with small numbers of classrooms as teachers may be identifiable indirectly by colleagues or parents, even if names are changed

(Taber, 2013). This is an issue, especially if teachers' practice is to be critiqued. Therefore, privacy was an ongoing concern for this research. This is particularly problematic for the stories of change, which provide a great deal of detail about each teacher's practice. This issue can be eased by seeking the teachers' approval and providing opportunities to review work. For example, when some initial (mostly positive) stories of change were published as part of a report on the research group's activities, the teachers concerned were given the opportunity to review the materials beforehand. However, because of the nature of this work – understanding the changes teachers were or were not able to make and why – the analyses were sometimes critical of teachers' practice, insofar as children did not have much autonomy or teachers' attempts were unsuccessful. Whilst this struggle was voiced by the teachers themselves, there is a risk that sharing a researcher's summaries, which would inevitably include the researcher's own interpretations, would create in the teacher feelings of vulnerability and loss of confidence. In addition, it would go against the precautions that were made to ensure classroom visits and interviews were not experienced as judgemental. This is not to say that the analyses are a judgement on the teachers, and I argue frequently throughout this thesis that what is seen as good practice is very much dependent on one's objectives and views about the purposes of education. Therefore, I was very careful not to label or judge teachers, and I have written about my own reflections in a piece for my Faculty's student association website (Le Courtois, 2019). However, it is a fine line between critical analysis and judgement, and teachers may interpret the analyses in a negative way. Therefore sharing personal, rather than summative, results may not always be in the teachers' own interest.

Authorship and teacher-collaborators. A related issue is that of authorship and the ambiguous role of the teachers as both participants and collaborators. During the programme, we emphasised to the teachers that they were partners in this inquiry: we were testing these ideas together and teachers were encouraged to adopt a research stance (Stenhouse, 1975) towards their practice. However, our approach was not as teacher-led as action research; the main focus was set by the research team and teachers did not participate in the more formal aspects of data collection and analysis. Through these analyses, I tell the teachers' stories, but through my own lens rather than giving them a voice. Whilst it could have been possible to have teachers write their own stories of change, this was not part of the design of the programme, mainly because of the time, training and increased commitment that this would have required. In addition, the programme was not set up as an action research programme but as a Community of Practice and teachers' involvement with the programme ended with the

final workshop. Finally, it would have resulted in a much more fragmented analysis, rather than an overall story that was nonetheless sensitive to context.

As a result, it is possible that the teachers still perceived themselves to be participants, rather than collaborators, in the programme. As described above, there were many aspects of the programme which had been predetermined and over which teachers did not have control. This may have resulted in a power imbalance between the research team and the teachers (Gore & Gitlin, 2004). For example, despite our intention to establish a dialogue with transfer of knowledge in both directions (McIntyre, 2005), teachers may still have perceived the programme as something mainly aiming to transfer research knowledge onto them. This is suggested in interviews where teachers referred to what they had gained and learned from taking part in the programme, but rarely mentioned what they had contributed to the researchers' understanding or to theory building. One exception is a teacher who explained: "You've got the research side ... we'll go and try it, and we'll tell you in the real world how it went!" (Stephanie, final interview). This suggests that this teacher perceived their participation in the programme not only as valuable to their own practice, but to the team's academic aims as well. It would be interesting in future work for us to assess the extent to which teachers had positioned themselves as knowledge and theory creators (Pascal & Bertram, 2012), and to make this role more explicit to them.

Chapter 4 - Contiguous Analysis: Teachers' Stories of Change Through the Stepping Stones Programme (R.Q. 1)

As well as common themes which we discuss in the next chapter, there were important differences in teachers' experiences of the programme and in their capacity to make changes to their practice. Not only were teachers' school contexts different, the programme encouraged teachers to set their own goals and teachers chose different strategies to experiment with in their classrooms (Appendix I). In addition, even when teachers chose the same strategies, they often interpreted and implemented them differently. To capture this diversity of trajectories, in this section, I present four contrasting cases as 'stories of change', describing the teachers' practice at the start of the programme as well as their attempts to provide children with greater autonomy. These stories were chosen because they provide a good representation of the wide array of experiences teachers reported and practices we observed: Lisa had a very constrained context, with very little autonomy of her own, whereas Julie and Beth both had supportive head teachers; both Anne and Lisa struggled with the practices promoted by the project, but for different reasons; both Julie and Beth embraced the project but the practices they focused on were very different.

These teachers were all highly experienced, each with around 10 years of experience in schools, sometimes more. Lesson observation points are referred to as 'initial', 'mid-point' and 'final' as per Figure 6 in Chapter 3, in reference to when they occurred in relation to the programme's start and end.

4.1 Anne's Story

4.1.1 Anne's General Teaching Approach Before the Programme

Anne taught part time: two days a week she was the teacher in a mixed Reception/Year 1 class and one afternoon a week she taught Science to a mixed Year 1/Year 2 class. Anne was a warm teacher and was enthusiastic about children's work and their ideas. At the same time, Anne was the one in control, deciding what the children did and learned: "I like to know what I'm doing and have a clear objective. I don't like it when it's all chaotic" (initial interview).

Anne saw the teacher as being a structuring force in the classroom and described this structure as being important for the children to learn: "I think they gain confidence from

having a clear framework, then they can move on with that” (initial interview). In both of her classrooms, children accessed new knowledge and practiced skills through whole class instruction and teacher-set structured tasks. Her younger class was typical of a Reception setting in England, with extended periods of time where children could choose activities that had been set out (a practice known as continuous provision or ‘free flow’). Whilst children were playing, Anne often worked with small groups of children on a more academic task and usually provided intense step-by-step support. The older class was much more formal, typical of Primary schools in England, with children usually seated on the carpet or at desks, and with more structured activities. The children had more limited time to do activities they had chosen for themselves.

Despite the importance Anne gave to structure and control, she gave the children some space for their own ideas. For example, in the first Science lesson we observed (initial lesson 1), which focused on the properties of materials, she provided open-ended activities in which children could be creative and which they became engaged in. Children could make shapes with playdough, make objects from twisting wire and find ways of folding a piece of paper so it could carry a weight. However, it seemed that these open-ended activities were distinct from the learning she wanted children to achieve. The children could only do these open-ended activities once they had finished a worksheet which Anne saw as “the main bit of the lesson” (initial lesson 1 debrief). In addition, during the open-ended activities, the children were rarely engaged in conversations about the properties of the materials they were using. For example, during the plenary Anne shared the surprising finding one child had made that playdough does not stretch, but she did not pursue this conversation further and did not use the opportunity to spark enquiries or encourage exploration. We did not record any other discussion of the properties of materials in the open-ended activities except Anne’s occasional use of science vocabulary (e.g. “You’ve really squashed and bent that, haven’t you?” initial lesson 1).

In the younger classroom, children had considerable choice over what to play with during continuous provision. Some activities were open-ended, such as a playhouse, but a large number were linked to her focus for the lesson and closed tasks. For example, after a whole class instruction on Easter, children could colour in crosses and write Easter-related words on them. Overall, children were not empowered to take control of the direction of the learning, which Anne felt should rest with the teacher.

4.1.2 Taking Part in the Programme

Balancing Freedom and Structure. Anne struggled with the ideas in the programme and attributed her struggle to her need to feel in control and have a clear direction for children's learning. Although Anne's first goal she set herself was to be more flexible in her planning in order to follow children's interests, she was not comfortable with the idea that children could be involved in making decisions about their learning: "The idea of it is they're dictating where they're going to end up. I don't know, something feels slightly... I think it's good but something slightly goes against the grain with it. So I just feel a bit uncomfortable with it." (final interview)

This struggle is evident in her mid-point science lesson with her older classroom (mid-point lesson 1), where she had planned for the children to generate questions about plants, then choose one question to do a fair test (an investigation where only one variable is changed, sometimes known as 'control of variables strategy'). However, two main issues arose. First, as she asked the children to generate questions, some of the children became involved in discussions. When a child asked, "Can plants grow without water?", other children shouted out answers and began debating (e.g. "But they don't have water in the desert"). Anne told the children that, although "debate is good", this was not what they were focusing on right now. She told them there would be a time for answering the questions "in a minute", but moved onto the fair test. As a result, she never gave them the chance to continue debating the answer and discuss how they might settle the debate or find their own answers. She was focused on her goal for the children to do a fair test, and she stopped their lively discussion.

The second issue arose during the fair test. She demonstrated how to do a fair test, then asked children to plan and conduct their own in small groups. However, many children's investigations were not fair tests, the room became quite noisy, and many children rushed their investigation to move on to other activities that were available once they were finished, such as painting. Anne described the lesson as "quite chaotic" (mid-point lesson 1 debrief), "a total disaster" (Community of Practice 2) and "out of control" (final interview). It reinforced her sense that giving children more freedom meant a loss of control: "I naturally I feel out of control with it. And that lesson was out of control. There was too much choice." (final interview)

Small changes. Despite struggling with the main thrust of the programme, Anne had taken some small ideas on board. Her whole class instructions to children became shorter and she gave children more space to talk about ideas. As part of the goals she has set herself in Workshop 1, Anne had chosen to try elaborating on children's answers, and increasingly asked questions to draw out children's ideas. For example, when a child showed her a catkin during a garden exploration activity, she asked the child questions which invited the child to explain he thought it was a flower because it "looks like petals" (mid-point lesson 2 running record).

The tasks she provided were also sometimes more open-ended than they had been before the programme, with greater space for children's ideas and interests. For example, in the final science lesson with the older class, the children were asked to do a worksheet about the body parts of invertebrates then to create an invertebrate out of a range of materials. They could choose what creature to make, what materials to use, and had freedom as to how to represent the creature as long as it had the correct number of body parts and legs. They also had laptops available to research information if they wished to.

Nevertheless, Anne remained the one in control of the purpose of the tasks and what she wanted the children to achieve – there was no conversation or involvement of the children in what they might create, how, and why, and Anne was not always open to children's own ideas and initiatives. For example, when a child showed her a flower he had painted (instead of an invertebrate), she replied: "That's a beautiful flower. Now I'd like you to do a scientific worm" (final lesson 1). Instead of using this child's interest (the previous term's topic) as an opportunity to extend his knowledge, she redirected him towards her lesson's objective, even telling him which invertebrate he should draw.

A Return to Previous Practice. In our final visit, Anne had reverted back to her original way of running lessons. Anne had also begun applying her idea of structure to the younger class. For example, after the last lesson we observed, she remarked that the children in the free flow setting had not chosen a focus activity she had prepared. She commented that she should have given the children a "clear definite task they all had to do" (final lesson 2 running record) to ensure all the children did this task rather than choose the one they wanted as usual.

4.1.3 Summary of Anne's Story

Anne struggled with the ideas in the programme, found them risky and uncomfortable. Although she was an active contributor to discussions during Community of Practice workshops, Anne's story does not suggest empowerment for her but a sense of discomfort. She felt a strong need for structure and control in her classroom, and therefore her attempts to try new practices made her uncertain about her practice, especially when she encountered difficulties. She made some small changes to the space she gave for discussions with children and for meaningful open-ended tasks, but her reflections and feelings of uncertainty led her at times to be even more controlling.

4.2 Julie's Story

4.2.1 Julie's General Teaching Approach Before the Programme

From 'Chalk and Talk' to Continuous Provision. Julie described her approach until recently as "the old-fashioned chalk and talk" (initial interview). However, the new head teacher at her school had asked Julie and her partner teacher to put into place continuous provision. In the school's version of continuous provision, which Julie described as "other learning through play alongside the formal lessons" (initial interview), the children were split into four groups and rotated around different areas over an extended period of time. In some areas, the children could freely choose activities including some play activities and some more academic activities whilst supported by a Teaching Assistant (TA). During that time, Julie worked with a small group of children on a pre-planned, structured task.

The timing of joining our research programme therefore coincided with this new development and the large change in practice it required from the teachers, which Julie described as "quite exciting" (initial interview), although she also reported that she and her partner teacher had received little support.

Choice and Autonomy. Compared to more formal classrooms, the children in Julie's classroom had much greater choice and autonomy, even before we began working with her. When not working directly with her, the children could choose from a large range of activities, as the children in Anne's younger class did. Julie also provided space for children's ideas and for their thinking to develop. For example, in one of the first lessons we observed, she told the small group she was working with to "think what you could write down to explain to someone how you solved this problem" (initial lesson 1). Julie also sometimes

dealt with ideas that did not fit her preconceptions by asking children to elaborate. For example, when asked to paint a picture about spring, a boy had painted a multicoloured picture. Julie asked him, “Can you tell us about your tree?” and the child explained that he had chosen “different colours for blossom and small leaves that are just opening” (initial lesson 2)

Julie also tried to help children solve problems on their own by asking them supportive questions. When a child asked if they could take their coat off outside, Julie replied, “That’s for you to choose, can you play with your coat on? Are you warm enough, then you can decide if you want to take your coat off” (initial lesson 1). She not only gave the child the control over the decision, she helped him think through it so he would make a choice he was happy with.

Julie’s Attitude to Continuous Provision. Julie was very positive about the use of continuous provision. Already before the programme started, she had seen benefits of continuous provision in terms of children’s engagement. However, some aspects were causing her some difficulties. For example, some children sometimes engaged in more frivolous play, which Julie did not feel resulted in learning. For example, in one of the initial lessons, three boys were playing in the pretend play area with books. Julie stepped in and asked one of the children, “What are you playing? Are you reading a book? Are you playing pretend?”. When the child shrugged, Julie “Think about what you are doing okay, remember to think about what you are playing”. She then pointed to two activities in the free flow area and directed him to the maths table with a string pin board (initial lesson 1).

Teacher Control in Formal Learning. Whilst children had many opportunities to choose activities in continuous provision (which is unusual in Year 1 in England), in many ways Julie’s classroom was like Anne’s younger classroom: the teacher’s role was to work closely on a set task, whilst the rest of the children play. Although less obviously than Anne, Julie also retained a clear agenda of what the lesson objective was for and steered children in that direction. The tasks she set were closed, with clear right answers and a specific way of doing them. In the first lesson we observed, she used a child’s book as an example of what to do. One child began to (correctly) write a list of numbers under the number grid in the workbook. Although the list was correct, it was not what Julie wanted, and so she told the child she was not doing the activity correctly because she “didn’t listen” (initial lesson 1).

In addition, Julie used time with her whole class in question-and-answer ('ping-pong') conversations, with the teacher firmly in control of who speaks and what questions are to be answered, rather than to generate discussions.

4.2.2 Taking Part in the Programme

Taking part in the programme gave Julie concrete tools she could put in place to realise some of the potential of children's autonomy, in particular in continuous provision.

Increasing Discussions. As the programme went on, Julie engaged the children in deeper and more open conversations where children could explore ideas. For example, in one of her mid-point lessons, she asked the children to think about what halving meant. She gave each pair of children a piece of paper and asked them to experiment different ways of halving it. She also asked them to justify their thinking, e.g. "How can you prove that they are the same?" (mid-point lesson 2) and to find more than one way if they could. Her questions in lessons often invited children's thoughts. For example, in a science lesson she asked children to discuss an experiment by asking, "Has it worked?" and "What can you see?" (mid-point lesson 1).

Flexible Planning. One of Julie's initial goals for the programme was to be flexible in her science planning by using children's questions and interests. Following a lesson where children generated hypotheses about a dead plant, Julie involved the children in discussions as to what to do next. She invited them to write their hypotheses on post-it notes so they could refer to them later. Importantly, she then followed up on these hypotheses by planning three further lessons based on them. For example, the children explored the outside area for clues of the Easter Bunny (whom they had accused of killing the plant). In another lesson, she gave children beans and soil and involved them in a conversation about how to help them grow. Each bean was then planted in a different way based on what the children thought would help it grow. The children also asked for a garden centre and to grow their own plants, so she integrated this into her continuous provision. She also let children follow their ideas through to their logical conclusions: she helped children investigate the idea that the Easter Bunny could have killed the plant, and when some children suggested giving a fruit drink to the bean so jelly beans would grow, Julie let them test the idea for themselves.

Open Tasks. Julie also increased the openness of her tasks. In her final Maths lesson, during continuous provision Julie encouraged children to explore numbers and patterns using resources in the classroom. One child began putting numbers on a number line. When Julie pointed out some numbers were missing, the child began adding them in using pieces of paper. Other children joined him as they decided to extend the number line to 100, working collaboratively. Julie checked in on them at different points in the lesson, providing suggestions without giving the solution, for example asking him to count for her and to explain how he filled the gaps:

The child stops at an error and the teacher asks, “Ah, have you noticed something wrong?”. He says, “These are the wrong way around”, and looks puzzled. The teacher looks and says, “Hmm ok, so some are missing and some are the wrong way around, so you might need to check the number line” (final lesson 2).

She also offered the children the chance to continue the task when they would have usually had to move on to the worksheet.

Providing Support. In order to support children’s increased freedom, Julie gave them tools and scaffolds to use independently. For example, in the Maths lesson on halving, children were working on blank pages in their workbook, cutting and sticking shapes into it, drawing around templates and writing captions but Julie had stuck some prompts in these books to guide them to use it this way (mid-point lesson 2). She also used questioning in a way that encouraged children to reflect, asking questions such as “What can you notice? What similarities can you see?” (mid-point lesson 1) rather than giving them the solution.

Retaining Teacher Control. Whilst Julie was positive about the programme and gave children increasing space for their ideas, children’s learning was still centred around teacher-imposed objectives and tasks.

Julie often spoke of children exploring and representing their thinking in different ways, but their autonomy over the learning content was not her focus. Her lessons retained objectives, though they were broad (e.g. halving), with always one set task the children had to do with her. In particular, in her Phonics lessons she set the pace and there were few opportunities for children to contribute to the shape of the lesson. As a result, she sometimes missed opportunities to follow children’s interests where they did not match her pre-existing plan. For example, in her mid-point lesson on plants, the children had made interesting

discoveries (in particular that one plant placed in the dark had grown a lot). Instead of encouraging further discussion or investigation, she moved on and continued with her plan for the lesson.

4.2.3 Summary of Julie's Story

Julie was positive about the programme, and in many ways showed she was actively taking control of the changes she was making. Julie had already begun implementing a large change to her practice, so she changed little over and above the shift to continuous provision. However, the programme gave her support and tools to use, and a community to reflect with on what she could improve. As a result, the spaces she gave for children's autonomy increased and became more refined over time.

4.3 Lisa's Story

4.3.1 Lisa's General Approach Before the Programme

A Focus on Compliance and 'Expected' Behaviour. Children's behaviour (in particular lack of engagement and disruptive behaviour) seemed to be an important concern for Lisa: "They've got to get that right before they can do the learning bit" (initial interview). She spoke of "cracking" behaviour and spent a lot of time in lessons managing behaviour so children were on task. She saw choice as relating to whether children complied or not: "I think the first thing, for this particular class, is what behaviour choice they make. That's the biggest choice they make" (initial interview). Lisa's school had recently introduced a new approach to behaviour management, which Lisa liked, and which emphasised expectations for behaviour. As a result, Lisa very often spoke to children about expected behaviour, for example calling down "3... 2... 1..." to show her "expectation" (initial lesson 1).

However, she also noted that the approach was focused on positive behaviour, for example by giving out positive behaviour notes. She also emphasised the importance of praise, which was very salient in her class but directed at compliant behaviour. She also sometimes used threats of sanctions, like losing playtime.

Prescribed and Scripted Lessons. Lisa's school had very strict requirements in terms of planning and timetabling. Children had set activities to do from the moment they walked into the classroom in the morning. Literacy and numeracy lessons focused heavily on teacher instruction, often at a fast pace, followed by independent practice. The literacy

lessons (which included phonics) were based on a scripted scheme of work and Maths lessons were planned using a flipchart template which was the same for the whole school. Lisa controlled the pace, and during independent activities she gave children frequent reminders of how long they had left.

Because these lessons were scripted and followed an imposed formula, Lisa had very little freedom to give children more autonomy within them, as she herself had none: “It’s very prescribed in what you teach and when. So, yes, you can push the children on in that little bit by saying ‘Can you add an adjective in there?’, but actually it’s not part of the script you’re meant to do” (initial interview).

In the lessons, there was a focus on memorisation, and often chanting or repeating of phrases. For example, the teacher asked children to repeat key learning points at different times during the lesson (e.g. “teen numbers... the one comes first”, initial lesson 1). Lisa saw these as important for children to remember the lesson.

Topic as the Exception. As an exception to the scripted lessons described above, one afternoon a week the children had ‘topic’ lessons – that is to say a more cross-curricular approach to learning which encompassed the humanities as well as aspects of the science curriculum. In the initial lesson we observed, the children were finishing a term’s work on continents and which animals lived in different parts of the world by making a display. The children had made labels for the continents and drawn animals to populate the display. While the task was set by Lisa, there was a purpose to it in creating a display. There was also a more relaxed atmosphere in the classroom as children could work at their own pace, could get up to get resources or bring their animal to Lisa. The children could also choose what animals they wanted to draw, though Lisa sometimes limited their options if too many children were drawing the same animal. As the children came to Lisa, she engaged them in conversations about where their animal lived, and elicited their knowledge about the continents and the oceans, or involved them in reading the labels.

4.3.2 Taking Part in the Programme

Elaboration. Elaboration was the first strategy Lisa was interested in trying. As a prompt, she had shown the children two pictures of cars, one of a toy car and one of a real car, and asked the children to compare them. She already had an objective for where she wanted the children’s discussion to lead, because she had planned a specific task to follow

from the discussion (“the outcome was that the children had to make a moving car via a balloon ... where I wanted to go [with the discussion] was to find out what it needed to do to make it move and about the wheels”, CI interview).

Therefore, although the prompt was open, Lisa had a very specific idea of what she wanted the children to discuss, and what conclusions she wanted them to reach. When some children began talking about other aspects of the cars, such as their colour, or other components needed for making them move, she closed off those conversations: “because I wanted to keep going on that wheel element, I said, Right we’ll park that idea, we’ll put that there and we’ll come back to it” (CI interview).

Lisa’s use of elaboration as a pre-scripted conversation towards a specific end point was different to the elaboration example we had given the teachers, where the purpose had been to elicit children’s ideas and to build children’s meaning making through the process of elaboration. For example, Lisa could also have been more open to divergent paths of discussion, because children’s comments on colour and the engine also revealed something about their understanding of what makes cars move and about the properties of materials (the national curriculum requirement) that she could have delved into.

Accepting a Range of Answers. In small ways, Lisa continued to try elaboration throughout the programme, by encouraging children to explain and evaluate answers more. In particular, Lisa explained she was trying this in numeracy: “It became more natural to do it in Maths. So I wouldn’t kind of accept one answer ... trying to get out of them a bit more information” (final interview). We did see some small moments where Lisa encouraged children to pursue their answers further:

The teacher picks another child for the next question. The child gives the wrong answer. The teacher writes the answer, then asks another child, then a third child, each time writing the number they gave. Then says, “Let’s see who’s right, let’s draw a picture.” The teacher draws circles, and asks children questions at each step of the process, e.g. “How many am I taking away [child’s name]?” (mid-point lesson 2)

In this example, we can see Lisa becoming a little more responsive to children’s ideas in the moment. Lisa’s emphasis for the lesson was on children using pictorial representations to work out arithmetic problems. When the child gave the wrong answer, she used this as an opportunity to show the value of using pictorial representations, because it could help the children decide “who was right”. However, she did not engage the children in conversations

about how they had done it themselves (therefore helping them work out where they had gone wrong) and instead modelled how to arrive at the correct answer.

Giving More Choices. In one of her topic lessons, Lisa wanted to give her children more choice. The task was to make an advertisement for a toy (based on their topic), but Lisa left it completely open to them how to make the advert. However, she found that a number of children “couldn’t cope” (Workshop 3): some used iPads to record themselves, but instead of making an advert were making noises; other children walked around or easily gave up. Lisa wondered whether it was because they were not used to the choice: “I don't know if it was because we take the element of choice away from them so they don't know how to respond in that situation” (Workshop3). She wondered whether they therefore needed more time and practice.

Lisa did not consider whether the choice had been a meaningful one and whether they had the right skills, not for making choices, but for the task itself. For example, she did not talk about whether the children had been interested in making an advert or whether it was part of a larger purpose. It may have been children did not know what they should include in an advert, or had not given much thought as to what makes an advert good or not. The task was decontextualised and Lisa still controlled what children achieved in the lesson – children could choose between options but were not active participants in the learning journey.

However, like Anne, rather than considering how to make the learning more meaningful, Lisa came to the conclusion that children’s choices needed to be limited: “It's good to give the children choice but in almost directed choice. So not give them a choice in everything because then they get really overwhelmed, but just give them a choice of a few things” (final interview).

Constrained Context. As Lisa emphasised many times, her context did not allow her much freedom to experiment with her practice, and her attempts were limited to one afternoon a week during topic. Some practices, in particular continuous provision, were off-limits because they required changes to the classroom organisation. Therefore, unsurprisingly Lisa’s Maths and Literacy lesson remained mostly unchanged. She continued to emphasise children’s non-compliance as choice (“Lots of people are choosing not to meet my expectations this morning”; mid-point lesson 2) and on using verbal injunctions, praise and sanctions to manage behaviour (“Oh dear who do I need to put on my list for missing a minute at lunchtime”; mid-point lesson 2). She also had very precise learning objectives for

each lesson, which she decided, and this meant children seemed to have little ownership over their learning, even in her attempts to use more elaboration and give more choices.

4.3.3 Summary of Lisa's Story

Overall the children in Lisa's class were in a very structured environment, with an emphasis on schedules, fast pace, didactic instruction and repetitive practice in order to acquire a high level of academic skills – practices which were largely imposed by her school. In such scripted lessons, children had almost no autonomy, and neither did Lisa. Through the programme, Lisa tried to draw children's ideas out more through elaboration. Whilst this was limited by her own scripts for the lessons and a focus on the correct answer, it was a conscious effort from Lisa to make more space for children's thinking. Lisa also attempted to give children more choice, but Lisa understood choice to be about having options rather than power. As she tightly controlled what tasks were and their purpose, children were not empowered to make the learning their own.

4.4 Beth's Story

4.4.1 Beth's General Approach Before the Programme

Teacher-Directed Lessons. Beth initially described her teaching style as “old school”, which she described as lessons “where there is less chaos”. However, she also stated she was “learning to appreciate chaos sometimes” and also stressed the importance of not sticking rigidly to a style, instead adapting to what different classes need (initial interview). The school followed a number of published schemes of work the school had purchased, but she also described a lot of flexibility within the timetable. Beth's lessons before the programme were very structured and with tightly controlled objectives: “we don't very often, in here anyway, have a choice of what they can do” (initial interview).

Whilst she was warm in her tone, she was often quite directive. For example, at the end of the first lesson we observed, children had to go for a run. Beth gave strict instructions to run two laps: “You are going to run, you're not dawdling ... No, you gotta run” (initial lesson 1).

Lessons were carefully structured around a typical starter-input-practice format. For example, in a Maths lesson, the children had to count up in different intervals as a pair. Then the teacher gave the learning objective (“Today you are going to carry on making arrays”, mid-point lesson 2) and taught children how to do this. Then the children had a period of

independent practice, whilst she walked around the classroom to help any child who needed support. In this very structured type of lesson, Beth decided almost everything: what children would learn, why, when and how they were going to learn it. However, children did have some autonomy and freedom in some areas.

Children Choosing What They Need. Beth noted the children often had choice in the methods they could use. For example, to do sums in Maths, children would be able to use concrete objects or to draw pictures in order to find the answers. This means that Beth encouraged children to be aware of what they needed in order to successfully complete tasks and allowed them the freedom to meet those needs themselves, rather than her deciding on children's behalf.

Children's choice over whether they used different supports for doing calculations also meant children could decide for themselves how abstractly they wanted to work – that is to say not just the methods, but how difficult the task was for them. In English lessons, she also sometimes gave the children choice over the level of difficulty by having one set of 'success criteria' for everyone ("This is what everybody has to have", initial interview), but would also include some additional 'challenges' for children to try. By giving these to everyone, rather than targeting them at children she wanted to push, she suggested it empowered children to challenge themselves: "they may not be able to grasp the basics of a full-stop, a capital letter, but they'll try their hardest and use a connective" (initial interview).

Making Learning Meaningful. Whilst Beth said she taught many lessons as 'stand-alone', focused on skills and knowledge children needed to acquire, she also spent time putting these skills into practice so that they would be relevant to the children, for example by doing inquiries or projects. For example, in a previous unit on 'life below water', the children made a sea life habitat inside a shoe box and had to create a narration in the style of David Attenborough in Blue Planet (a nature programme). In preparation, the class watched clips from Blue Planet, as well as videos from their scheme of work on presentation skills. The learning about habitats was also integrated into the project, as children needed this knowledge to create the shoe box. Whilst the children did not have choice in what the project was, the knowledge and skills they acquired were made relevant because they were learned for a purpose.

Beth's Buddy System. In addition to being able to make choices about their preferred ways of working, and to have opportunities for purposeful learning, the children in Beth's class had some choice over who they worked with, through what Beth called a 'buddy system'. At the start of each week, Beth picked children at random and they could choose the person they would work with for the rest of that week – their buddy. They then sat with that person in all lessons, and whenever partner work was required, their buddy would be their partner. Though Beth's system for allowing children to choose partners was very structured and therefore had little flexibility (for example, children could not change their mind, or decide that for a specific task another child would be a better partner), it nonetheless allowed the children to experience some say over who they worked with, and therefore could learn the value of working with different people without this being imposed on them.

4.4.2 Taking Part in the Programme

Using Children's Ideas for Planning. Although the first goal she set herself was elaboration, Beth's main focus during the programme was flexible planning in Science lessons (which happened once weekly), in particular basing her medium-term planning on what the children were interested in ("We went for no formal planning", final interview). She used the first lesson to stimulate children's interest and curiosity, and then planned the next lessons based on what questions emerged in this and further lessons.

However, whilst the planning was very child-led, Beth was also conscious of having to meet certain objectives. She described the process of following children's lead whilst also meeting National Curriculum requirements as "weaving". Key to this process was Beth's flexibility, as she kept it open as to where the children could take the learning, planning in response to what had happened during lessons and changing her plans when needed: "Although I've had an idea going forward week by week, if they've thrown up something, we've changed it" (final interview).

In order to kick start the inquiry, in the first lesson she took the class to the school grounds, where there was an allotment and a field. Beth asked the children to collect some dead plants (as this was the hook she had agreed with other teachers in the Community of Practice), and back in the classroom she asked the children to observe the plants, which led to children asking questions about them. Beth made a note of the questions the children generated in order to plan her next lessons, based on what the children had expressed interest in and had been curious about.

After the initial question-asking session, Beth planned a couple of lessons to begin answering some of these questions, including an experiment to grow rocket in different experimental conditions over a couple of weeks. This created further interest and discussions, which Beth then used to plan a further set of lessons. When this particular line of inquiry was exhausted, she chose from another set of questions that some children had asked and which she knew would fit well with the curriculum. This led them to spend some time learning about the different kinds of plants and the parts of the plant. Another child had hypothesised that cut flowers would live forever because they had everything they needed, so Beth bought “a cheap bunch of carnation” and left them on the windowsill. During one lesson, the children made observational sketching of the fresh flowers, and Beth had planned to repeat the exercise four weeks later, “when they're nice and crispy and to look at the comparison there” (final interview).

Generating Genuine Discussion: Curiosity and Space for Experimenting. By taking children’s questions as the basis for her planning, Beth did not just follow children’s interests, but generated learning opportunities which mattered to the children. The rocket experiment was not only significant because it had been based on children’s ideas (in other words, an example of flexible planning), it also allowed the children to see themselves whether their theories were correct as to what had caused the plant to die, and therefore set the stage for genuine conversations to make sense of the world.

For example, some children had hypothesised that the plant had died because it had been left in the dark (based on prior knowledge that plants need light in order to grow). As part of the experiment, other children had wanted to try growing the rocket in the sand, or to be watered with orange juice, or to be grown in marshmallows. They were fascinated when the results were not what they expected: watering the marshmallow-grown rocket made mould, and the rocket grew both in the dark and in the sand. These seemingly strange experiments also gave the teacher a unique opportunity to assess the children’s understanding around plants: that marshmallows give energy because they’re full of sugar, that only seaweed can grow in sand, etc.

Some Limitations. Not all lessons were inquiries or as closely tied to children’s interests. For example, later on in the term, the children were tasked with making 3D plants as part of a DT lesson using “some plastic cups and pipe cleaners and cake cases” in order to produce photographs to be labelled in their books as evidence. In another lesson, Beth

focused the learning on different types of trees, which was not something the children had expressed interest in but which she needed to cover to meet National Curriculum requirement. Also, the class only investigated a small number of questions over the course of the term, and therefore some uncertainty remained around individual children's feeling of ownership if their question had not been picked.

Finally, Beth's fruitful trial of flexible planning was completely limited to her science lessons – in other words, to about an hour each week – whilst the rest of her teaching practice remained more or less unchanged. For example, in her final writing lesson, Beth was still very precise about what the objective was for the lesson, and what children should achieve. However, she did feel that it had changed her attitude to letting children guide her teaching and that this was having ripples in other subjects. For example, she noted that in Maths, it had made her “more open in a lesson to stopping and being flexible within that lesson as to what actually the outcome of the lesson is” (final interview). She explained she might be more willing to deviate from her plans, for example by providing different materials than what she had intended or being responsive to children's suggestions.

4.4.3 Summary of Beth's Story

Whilst the programme was a radically different approach for Beth, from the beginning she was committed to following children's lead in Science lessons, letting go of any preconceived plan. These genuine opportunities to explore and pursue lines of inquiry children were interested in meant that children were engaged and active participants. Not only this, but Beth carefully supported children's inquiries through planning activities to help them answer their questions (such as sketching, or growing rockets in different conditions) and by taking the time to facilitate discussions based on children's ideas. Although she found the process sometimes “nerve racking” (final interview), she emphasised the children's increased interest and how much more meaningful it had made her lessons.

4.5 Conclusion to the Stories of Change

The stories of change present four contrasting trajectories of teachers in our programme. The teachers had different starting points (Julie had already begun implementing ‘continuous provision’ with many opportunities for autonomy, whilst children had little autonomy in other classrooms), as well as different contexts (Beth's head teacher gave her flexibility in her timetable, whereas Lisa was required to follow scripts). They also had

different personalities, different children in their classrooms, and different levels of interest in the practices we were encouraging – and this was the case across all the whole group of teachers taking part in the programme, rather than specific to the four stories we chose. These differences meant that the teachers engaged differently in the programme: they did not all try the same strategies, and did not implement them in the same way. Indeed, there was no pattern in the Q-sort as to which practices teachers found easier or more difficult to implement, as their opinions varied widely. Overall, none of the teachers in the Stepping Stones programme made large, transformative changes to their practice and teachers generally struggled with giving children greater control over their learning. However, the flexible approach to the programme meant that all teachers were able to try some strategies. Despite these differences, it is also possible to see common threads between these stories, and to explain what accounted for these differences as well as teachers' struggle with giving children more autonomy. This is the aim of the next chapter.

Chapter 5 - Thematic Analysis: Pockets of Space for Children's Autonomy in an Ecosystem of Teacher Control (R.Q. 1 continued)

Because love is an act of courage, not of fear, love is commitment to others. No matter where the oppressed are found, the act of love is commitment to their cause—the cause of liberation. (Paulo Freire, 1972/ 2000, p. 89)

5.1 Overview of Chapter 5

The aim of this chapter is to explain overall patterns in the changes the teachers made as part of the Stepping Stones programme, what happened when the teachers made those changes, and what affected their ability to make these changes. Whilst the specific contexts varied between teachers, I found two broad commonalities to the classroom context:

- classrooms were tightly controlled by the teacher, who functioned as the centre of power and knowledge. A network of beliefs, interacting practices and institutional pressures upheld this way of functioning, as part of a mutually reinforcing system, which I refer to as the 'teacher control ecosystem'.
- in contrast to this ecosystem of teacher control, and *at the same time*, teachers opened up spaces for children's active sense-making and ownership of learning, to varying degrees of success. However, whilst this happened in parallel, this was not an additional ecosystem that ran concurrently, that is to say a set of coherent interconnected practices (as would be found in certain pedagogies). Rather, space for children's autonomy happened through discrete practices that meshed more or less well with the rest of that teacher's practice, and sometimes conflicted with it.

Therefore, whilst the 'ecosystem of teacher control' was dominant, these contrasting ways of working were co-occurring, even before the programme began – teachers did not adopt one approach or the other, but were engaged in a process of negotiation. The programme therefore encouraged teachers and gave them permission to open up more spaces for children's autonomy. This process and its result were not uniform across schools, as seen in the stories of change. Teachers' ability to give more ownership of the learning to the children varied hugely between teachers, both before the programme started and as a result of the programme. How much ownership children had, and in what ways, depended on a number of other factors, which I also outline in this chapter.

This chapter, therefore, is less concerned with which strategies the teachers tried to implement from the Stepping Stones programme – and whether they ‘worked’ – than with the bigger question of how teachers can increase children’s autonomy in the classroom. As explained in the methodology, this analysis therefore moves away from discussing specific components of the programme. Instead, though the experiences of the teachers in the programme, I attempt to deepen our understanding of what providing children with opportunities for autonomy is like in practice, and what supports or impedes teachers’ efforts to do so.

5.2 The Teacher as the Centre of Power and Knowledge in the Classroom

In all classrooms we visited, the overarching *modus operandi* of the classroom was one where the teacher was in charge of most learning decisions, in particular with regards to what tasks children did, and where many of children’s movements and interactions were dictated by the teacher. Whilst this has implications for children’s autonomy and inner motivation, and for their freedom and empowerment in the classroom, the value of this very much depends on the goals of education one endorses – though as I have argued (Chapter 2) there is a moral justification to paying attention to children’s autonomy. Nonetheless, the intention with the analysis below is not to denounce these practices as *bad practice*. Based on my own experiences working and researching in Primary schools in England, I believe the practices described below, though varied, are the norm in most schools in England. In addition, the teachers in the programme were all skilled professionals who clearly loved their students, and it is not the aim here to criticise them.

Nonetheless, it is impossible to discuss questions around children’s empowerment, autonomy and inner motivation – and therefore how difficult a task teachers were accomplishing through the programme – without also understanding the extent to which children’s lives in classrooms were controlled by their teacher, and how this was in turn shaped by forces outside the classroom.

5.2.1 The Teacher’s Control Over the Classroom

Teacher Control of the Learning. In all lessons we observed, teachers had a specific learning objective which they wanted to reach by the end of the lesson and in all but one school this was a school requirement. In all but rare cases, the teachers indicated that this was the norm, and that the learning objective was decided by the teacher in her planning in

advance of the lesson itself. When asked what their intention was for a lesson during debriefs, teachers often referred to their learning objective.

Even when the learning objective was built on or taking into account children's interests (as we saw with Beth's story of change), it was a decision made exclusively by the teacher, in advance of the lesson, and it was usually fixed. Alongside the learning objective, the teachers also usually decided what they wanted children to achieve in the lesson and what their learning would look like. Sometimes this was formally introduced at the start of the lesson through 'success criteria', and when this happened these were always given by the teacher – we did not observe these being jointly agreed with children. For example, in one lesson Beth wrote a list of criteria for children's writing: that it should make sense and that children should use "finger space", capital letters and full stops, as well as a number of conjunctions (and, so, but, because).

The Teacher as Judge. Along with this, the teacher was often the judge of children's work: the teachers often were the ones to decide when work had met expectations, and we frequently observed children coming to see the teacher for approval that their work was finished or correct, or conversely sent back to their seat to do more work. This sent a strong implicit message that the teacher was the one who ultimately decided the quality of the work. For example, in her lesson on animals in different continents, Lisa rejected work that did not meet her expectations: "This isn't your neatest writing so it can't go on the board" (Lisa, initial lesson 2). Sometimes, this led teachers to describe children's work as "not good enough" (e.g. Vicky mid-point lesson 1) or to be critical of it ("Not overly impressed I have to say, not your best work", Beth, final lesson 1). They also sometimes suggested that children's work or behaviour needed to please the teacher: "If I don't see finger spaces in your writing, I'm going to be grumpy" (Beth, final lesson 1).

Teacher Control Over the Task and Pace. Similarly, in most lessons we observed, the teacher decided what tasks the children did (though, as I describe later, there was sometimes some choice between options or within the task). Especially in Maths and Literacy, tasks were often closed in that there was a specific outcome the children needed to have produced. Even when there was openness in how the children might reach that end point (for example in making sock puppets), it was rare that the end point itself was open ended.

Teachers also often directed the pace and sequence of activities in the lessons. The general format of lessons started with a whole class introduction (the *input*), usually on the

carpet, which varied from 10 to 30 minutes; this included the teacher introducing any new content they wanted the children to learn, including modelling how to do something. It might also include some individual or paired practice on white boards but for only a few minutes at a time, and the teacher tightly controlled the pace. The children were often solicited to contribute ideas and information, so that teachers were rarely lecturing but rather engaged in a 'question-and-answer' dynamic with the children:

The teacher talks to them about doubles/halves i.e., "if I gave you a sweetie but I wanted to give you double, what would I have to do?" Children give answers i.e. "eat it", "put it behind your back", "give me another sweetie". With the final answer, the teacher draws two sweets on the board. (Stephanie, mid-point lesson 1).

This was followed by a period of independent practice through a task set by the teacher. In general, except in continuous provision or carousel settings, all the children did the same activity at the same time, though their version of the task might be different to cater for different 'abilities'. In other words, at any given point in time, other than in continuous provision, children were rarely engaged in a task they had chosen for themselves.

Whilst children could work at their own pace on these set tasks, in many classrooms there was a time pressure to complete the task or a number of tasks. Teachers sometimes gave reminders of how long was left and how much they expected children to have done. Sometimes teachers moved children on, regardless of whether the child wanted to continue, because the whole class was working on a new task (e.g. "We're now thinking about our favourite question, and we are not doing any more writing", Vicky, mid-point lesson 1).

As the teachers decided what task children did in order to achieve the set learning for the lesson, the teachers also often decided the level of difficulty of the task through what is known as 'differentiation' – that is to say, that *different* children receive *different* instruction and work based on what they need to learn. This was usually done by having tasks of different levels of difficulty – commonly three. For example, in one of Lisa's Maths lessons, the three levels of difficulty were writing numbers 0-20, then 20-50 and 50-100. Teachers also often decided how much work children should do.

Teacher Control of Children's Bodies. Children's bodies – their movements, where they sat, how they sat – were also often under the teachers' control. About half of the classrooms seemed to have fixed seating plans – that is to say the children sat in the same place every day. Teachers also sometimes changed where children sat, either because of

disruptive behaviour or because they wanted children to work with specific people (e.g. Lisa, initial lesson; Beth, mid-point lesson 1). Many teachers also asked children to sit in particular ways, and in particularly asking children to sit upright.

In many classrooms, children were not allowed to move freely. When children were allowed to move around, this was usually for specific purposes, like getting materials. For example, in her classroom, Beth often designated specific children to help her hand out resources, and they were not allowed to leave their seat unless authorised to. On an occasion where the children got up without being told, Beth answered: “Don’t be silly, stay in your seat” (Beth, mid-point lesson 1).

However, how tightly controlled children’s movements were varied a lot between classrooms – for example, in Helen’s classroom, children were often able to choose where they sat as long as they were working with children they could work well with. How tight the control was also varied between lessons – children were often observed standing at their table (rather than sitting) during topic lessons, in particular if they were doing something manual like cutting or sewing, even in classrooms where children’s movements were much more regulated like in Lisa’s class.

Teacher Control of Talk and Interactions. The teacher also controlled much of the talk and interactions in the classroom – in other words, children were often only allowed to speak when they had been told to by the teacher, and on the topic the teacher had chosen. In whole class situations, teachers often chose which children answered the teacher’s questions. Some teachers used the ‘lolly stick’ method, where they draw children’s names (written on lolly sticks) at random to answer questions, whilst others picked children with their hand up to answer. The teacher also sometimes asked children to talk to their partner (the child sitting next to them) to discuss a question or to work with them on a task.

In practice, children were regularly observed speaking out of turn, either ‘calling out’ in whole group sessions, or engaging in off-task talk. However, in many classrooms this kind of talk was reprimanded by the teacher and calling out was sometimes ignored as a way of showing it was not the correct way to share ideas.

Through choosing who spoke, when and on what topic, the teacher controlled the flow of the conversation and what they wanted children to learn. Opportunities for children to talk needed to be part of the teacher’s plan, and as one teacher explained, talk was also a behaviour management tool (“It stops them from being so chatty ... it’s a good way to get the chattiness out, but in a productive way”, Claire, final lesson 1 debrief).

Whilst elaboration was one of the strategies some teachers tried, the way turns of talk occurred rarely resulted in deep meaning making. For example, in Anne's mid-point lesson, whilst Anne asked one or two follow-up questions, she did not explore children's ideas further than in a superficial way:

The teacher reminds the class about a child's idea to put the plant in the fridge, and writes on a flipchart "Can plants grow in the fridge?". Some children shout out "no", some say "It depends". The teacher asks, "What does it depend on?" and the child answers that it depends on the type of plant. The teacher then asks another question about plants. (Anne, mid-point lesson 1).

Instead of asking what the child meant by 'it depends on the type of plant', and using this to draw out a conversation between children about what plants need, and the fact that different plants might need different things, she moved on, because her aim was to generate questions, and answering the questions was not part of her plan. It shows how the conversation was controlled by the teacher and her agenda, through her allowing certain children to talk at specific times and not others, through the questions she asked and the ones she did not.

Children's interactions were also controlled through who children worked with. In some classrooms, children could choose their partner (either informally, or more formally like with Beth's buddy system) some of the time, but in other classrooms (e.g. Vicky, Lisa) children had a set place to sit, and this determined who they worked with.

5.2.2 Directive Language

Teachers frequently reinforced the message that they were the ones in the position of authority. They often spoke of what they wanted, and used direct commands to give instructions: e.g. "All I want you to do is write," Beth, final lesson 1; "Look at this, don't write on there yet," Vicky, mid-point lesson 2).

As seen in her story of change, Lisa frequently spoke to her class of her expectations for specific behaviours and achievements. Sometimes this position as decision maker was invoked to deal with issues. For example, in one of Vicky's lesson, a child had cut the string that was holding rockets the class had made while they dried, resulting in the rockets to tumble down. When reprimanding him, Vicky said: "No one has asked you to do that, very silly" (initial lesson 2), as though it was the fact that the child had acted out of his own

initiative, and not the consequences of his actions (to cause the rockets to fall), that was the issue.

5.2.3 Following the Lesson's Script

Controlling the lesson, such as talk and tasks, allowed the teacher to follow the plan she had made and produce the outcomes she wanted. In some cases (e.g. Lisa), there was a literal script the teacher needed to follow, but even more loosely planned lessons followed a pre-determined track to which teachers worked hard to keep. However, this also resulted in rigidity which sometimes closed off children's initiative and ideas. Teachers often did not follow children's initiatives or interests because they would lead them 'off track' and to run out of time. For example, Helen explained that her children had expressed an interest in some Indian clothes an adult had brought in (in relation to their topic), but that she had not been able to let the children investigate the clothes and try them on because she needed to move onto her Maths lesson (Helen, final interview). This also led teachers to sometimes dismiss what did not fit their preconceived idea of what children should answer. Teachers also frequently redirected conversations to be on topic. Sometimes, this was done subtly. For example, in one lesson, a child made a comment irrelevant to the learning objective – about a mouse he had seen at break time – but the teacher made links to the learning objective nonetheless to steer the conversation back on track. Often, teachers redirected conversations more explicitly, for example acknowledging that it was interesting but not what they would do now (e.g. Stephanie, final lesson 2). In debriefs, teachers sometimes mentioned these redirections were intentional, because they considered these discussions to be tangential to their plan: "We could have gone off on a tangent but we didn't. Actually, there was no need to" (Helen, final lesson 2 debrief).

The finding that teachers try to stick to their plan, and as a result close off initiatives or questions, echoes research in other settings, where teachers were described as shutting down children's questions in order to stay on script, even when those questions were relevant to their broader learning (Engel, 2011; Eshach et al., 2014).

5.2.4 A Focus on Correct Answers and Products

In many of the lessons we observed, the focus was on accuracy (e.g. correct answers) or the quality of the final product (e.g. a piece of writing, a sock puppet), rather than on the process – the experience, the sense making and the growth that might have been happening

through the activities. This means that there was often little space for children to make sense of ideas, and instead the purpose was for children to accumulate knowledge. As described above, teachers' feedback was mostly concerned with correcting children's work. For example, when reading a book with her group, Lisa's emphasis was on children's ability to locate the answer in the book (a test-passing skill), rather than on deeper understanding. This focus on the correct answer was also visible in classrooms discussions, where teachers were attempting to elicit children's ideas, but were focused on extracting the correct answer:

The teacher shows a picture of a beetle and asks, "Does anyone know what this is?" A child says, "Blue beetle." The teacher says it is a kind of beetle, and asks how many legs it would have. A child says "six" – the teacher asks, "If it has six legs, what does that mean it is?" She repeats the question with a prompt: "If it has six legs what is it, an in..." Children say "insect". (Danielle, final lesson 3).

This also happened when children did not give the correct answer, where teacher's questions were used to indicate to children they needed to reconsider their answer, though this also usually involved scaffolding to help the child reach the correct answer:

A child has written that a circle has zero sides and the teacher asks, "does a circle have zero sides? How many straight sides does it have?" the child says, "Zero", the teacher asks, "and how many sides does it have?", the child says, "One"... The teacher says, "Ah it has one!" (Stephanie, final lesson 1).

5.2.5 Reliance on Compliance and Coercion.

That teachers tightly controlled the classrooms and the learning does not mean that children necessarily experienced this as controlling or authoritarian. As we did not directly look at the impact different practices had on children, it is difficult to tell, though children's responses to some practices suggest they were passive in their learning (see section 5.5). One sign that children did experience teachers' control as going against their own will was that teachers frequently used compliance and coercion tactics, suggesting they needed to impose this control onto children.

In some classrooms, some children disrupted lessons; sometimes this was because of conflicts with other children or because they were not listening to each other's ideas (which made discussions difficult), but often it was because of their refusal to comply with teachers' requests or because of disengagement. Sometimes this resulted in defiance from the children, and direct requests by the teachers to comply. Children were also frequently observed off-

task (e.g. play fighting, singing), fidgeting, chatting or procrastinating (doing their work but very slowly), which suggests they were not particularly invested in their learning. Children sometimes rushed to get tasks done, or chose the easier option to avoid putting effort into their work, suggesting a lack of motivation.

Therefore, teachers often checked children were on-task, and used injunctions to remind them to focus (e.g. “Sit down like you've been asked”, Anne, final lesson 2). In some cases, children’s disengagement was such that they needed constant refocusing and one-to-one support to complete their task, though this may have also been partly due to other educational needs. In addition, teachers’ injunctions for maintaining children’s focus were not always directive and controlling.

Teachers also often mentioned that they needed to use strategies to keep children engaged and on task. For example, Claire mentioned that she sometimes used cold calling to keep children alert, “otherwise I find, with my class, they’ll just switch off, they’ll daydream, they’ll go. So, it’s kind of keeping them on the ball” (Claire, initial interview). Many teachers used praise to elicit the behaviours they wanted to see (though not exclusively for this purpose). For example, they praised children who were sitting quietly and listening, saying “thank you” for behaviour they wanted to see or giving positive remarks (e.g. “Excellent, Lottie, speedy quick”, Lisa, mid-point lesson 1; “You’ve sat really quietly this morning”, Beth, initial lesson 1).

This was highly effective, because the children wanted the teacher’s praise and attention. Teachers understood this need for recognition, and indeed often celebrated children’s work to help children feel valued, but also used it to produce desired behaviours. For example, Claire noted that evaluating children’s work as a class was “a really nice thing to do” because “when they hear the teacher celebrating this piece of work, it’s amazing how many of them will then go and produce that same piece of work to get that recognition” (Claire, final interview).

A number of teachers also used rewards to produce compliant behaviour. For example, Vicky used stickers to reward children who had done a lot of work (mid-point lesson 1). Lisa had a recognition board for children who were compliant, for example getting their maths resources and waiting quietly (initial lesson 1). Danielle had a clock with “golden minutes”, which were minutes of play time on Friday afternoons, and which children earned for the whole class. Some teachers also used sanctions (and threats of sanctions) to force children to comply with their requests. For example, Danielle had a visual traffic light

system, whereby children went up or down depending on their behaviour. Stephanie had a similar board, and emphasised to the children why children were in different areas:

She highlights children who were concentrated throughout, came up with new actions, joined in and were sensible. On the other hand, she highlights children who didn't respect the space of other children and didn't try to join in." (Stephanie, initial lesson 2)

Other teachers (Vicky, Anne, Lisa) sometimes threatened children with sanctions such as losing playtime. Whilst it is common for schools to use rewards and sanctions to produce desired behaviours, this behaviourist approach denotes a classroom culture focused on compliance rather than autonomy. It is also in contradiction with a focus on love of learning and inner motivation, as evidence suggests salient rewards undermine intrinsic motivation (e.g. Cerasoli et al., 2014; Deci et al., 2001). In settings where children's autonomy is valued, for example in democratic settings or in schools that use restorative approaches, issues such as disruptive behaviour and the use of sanctions are resolved through discussions involving the children themselves, rather than being in the sole control of the adults (e.g. Bevington, 2015; Hantzopoulos, 2013; Harrison, 2007)

5.2.6 The Teacher as Only Source of Knowledge and Help

The teacher not only had most of the power, but they also acted as the centre of knowledge. Whilst it may be understandable that teachers would be an important source of knowledge in the classroom, teachers (and TAs) were often the only source of information or help. We observed few instances of peer help and cooperation. This is perhaps unsurprising as children were usually engaged in individual tasks that required little cooperation. This is in contrast to pedagogies where peer learning is central to the way in which children access new knowledge, such as Reggio Emilia (Kim & Darling, 2009; McNally & Slutsky, 2017) and self-directed settings (Gray, 2017). One exception was in continuous provision, where children were seen cooperating on a number of occasions – I describe this further in section 5.4.

There were also few resources with which children could access learning without the teacher. There were sometimes practical resources children could use, such as cubes, to help them in their work, but if children lacked knowledge or information, the adults were usually their only source of help. That teachers were the only source of knowledge potentially explains teachers' reliance on whole class instruction: it allowed them to provide some input

to all children in an efficient manner. However, this efficiency came at a cost, as opportunities for children's autonomy in whole class instruction focused on knowledge acquisition was, as we've seen, minimal.

Teachers' positions as holders of knowledge perhaps also explains why lessons were so often focused on correct answers, rather than on children's own knowledge and meaning-making. Indeed, this suggests the ecosystem of teacher control was a web of practices that were inter-related and consistent with each other, creating reinforcing feedbacks.

5.3 The Broader Context: The Ecosystem of Teacher Control Beyond the Classroom

To understand why controlling practices were so prevalent and interwoven, it is important to place these practices within their broader institutional context. Here, I argue that teachers' need for control arose out of a culture of 'learning objectives', with three principal sources: high stakes accountability; content-focused interpretations of curricula; and direct pressure from top-down directives. All three sources can be seen as the result of a banking model of education (Freire, 1972/2000) with a focus on measurable academic knowledge, itself part of a wider international movement that views education through the lens of international rankings and test scores.

5.3.1 The Need for Control

Teachers' total control over the classroom is seen as good practice by some educationalists (e.g. Lemov, 2012). For many of the teachers in the programme, and in particular those who struggled with giving children greater autonomy, teachers felt strongly that they needed to be in control, and in particular in control of what children learned. A number of teachers explicitly mentioned the need to be in control as something that stopped them from giving children more autonomy. In her attempts to give children more control over their learning, Anne described her lessons as "out of control" and "chaos". She said, "I feel like I need to know where I'm going with things" and described being flexible as "brave" (Anne, final interview). She needed structure in order to feel confident in her teaching. She also described the idea of children having more freedom making her feel like they were "dictating" the lesson, which she did not like (Anne, final interview).

Other teachers also emphasised the need to be in control of the lesson and the class. Helen spoke of being "in charge of things" and being used to saying, "this is the lesson we're going to do" (Helen, final interview). She also noted that, before the programme, she

“certainly didn’t like the idea of just letting them have lots of choices” (Helen, final interview). Other teachers also echoed these feelings. Lisa described open tasks as not very successful because “the control of the class went” (Lisa, final interview).

However, this need to be in control was not just driven by teachers’ own ideas about children having power. It was also directly under pressure from school leadership and the influence of the broader education system. These institutional pressures often stopped teachers from following children’s lead and giving them more freedom, even when they wanted to.

5.3.2 The Learning Objective Mindset

Learning as Linear and Sequential Accumulation. The ecosystem of teacher control was driven by the idea that children need to learn specific things, and that the best way for children to learn them was for teachers to plan lessons targeting specific skills or knowledge to be acquired (the learning objective), building these lessons sequentially over time. Teachers’ responsibility was to build up this knowledge and fill any gaps that may arise. Whilst it is outside the scope of this thesis to show how this view has permeated education policy in England (see for example Fielding, 2001; Nicholl & McLellan, 2008; Wood, 2019), it is obvious in key documents such as the government-imposed teachers’ standards and frameworks from the Office of Standards in Education (i.e. the body responsible for inspecting schools and enforcing standards in England, known as Ofsted) (DfE, 2011; OFSTED, 2019b). In particular, these documents emphasise a focus on sequenced knowledge and progression, as well as a reductive definition of learning as “a change in long-term memory” (OFSTED, 2019a, p. 19).

This idea was also voiced by the teachers in our programme. Teachers had (as they are expected to) clear ideas of what they wanted children to learn and how they were going to build knowledge from one lesson to another. Beth, speaking of a scheme of work for Maths, explained: “it breaks down all of the elements that need teaching in the year into smaller steps and you follow them sequentially to build up the children's knowledge” (initial lesson 1 debrief). Helen explicitly mentioned this need for learning to be sequential, for some subjects (like Maths and Phonics) to be taught in a specific sequence in a specific way as a reason why it was impossible for children to lead the learning in those subjects: “They can’t lead you on something like phonics because they don’t know the next sounds they need to learn.

They don't know how to do the next bit of Maths or they'd have already been doing it" (Helen, final interview).

This mindset – that only the teacher could know what children needed to learn and how to learn it best – justified having a strict, pre-determined learning objective for each lesson, and meant that deviating from this learning objective was seen as problematic.

Keeping to the Learning Objective. The learning objective mindset was a major reason teachers taught in the way they did. Indeed, teachers' need for control over the classroom was largely driven by the need to meet the learning objective by the end of the lesson. Anne remarked, "half the problem's with an objective-based culture" which she called "stifling sometimes" (Anne, Workshop 3). Claire described how children had not carried out a task properly, by describing a picture rather than writing what the characters were saying (her own objective). She described how she worked to redirect them to it:

They brought their work to me and I said, so okay, you've told me what they're doing so now what could they be saying? So then they still met [the learning objective] by the end. (Claire, final lesson debrief)

The learning objective was also often the one thing that was non-negotiable. When teachers were able to give more space to children, this was always under the condition that children would still meet the learning objective. By contrast, teachers described how there was more space for children to lead their own learning when teachers stepped away from a tight learning objective.

5.3.3 High Stakes Accountability

The high stakes accountability system found in England (Alexander, 2011; Hutchings, 2015; Sturrock, 2021) meant that teachers needed to make sure their students performed well on tests and met certain standards, in particular because of the Phonics screening check at the end of Year 1, and the SATs national examinations at the end of Year 2 (which many Year 1 children were already preparing for). On a number of occasions, children were observed preparing for these tests, either through practice tests or through a focus on the content of the tests. Teachers described these tests and the data associated with them as sources of pressure because of "where the children are supposed to get to by the end of Year 1" (Danielle, final interview). Because the pressure was felt more in Literacy and

Maths, teachers felt it was more difficult to try the Stepping Stones programme strategies there: “the requirements for what’s expected that sort of get in the way a bit more” (Stephanie, final interview).

Some teachers mentioned this pressure went against their own wishes for teaching. Vicky felt tests were “not a true reflection on the character of the child” and it made her “angry” to be forced to go through them regardless: “we have to do it because we’re told we have to do it” (Vicky, initial interview).

Ofsted was also an important source of pressure that stopped teachers from being able to experiment with their practice. This high level of scrutiny meant teachers were often preoccupied with providing evidence and ‘data’ to show their students were making the expected progress. Helen talked about the need to produce work for accountability reasons, including by Ofsted, and for that work to look consistent, which constrained her ability to follow the children’s lead: “there’s so many constraints around what you have to do ... there is an expectation on what you’re teaching in terms of outcomes, but also what it looks like in books, the evidence trail even for topic, they expect things to look a certain way” (Helen, final interview).

It also meant that some teachers felt they were too busy with assessing children to have time for engaging with them in their play, to follow their lead or model ways of playing or helping them learn self-regulation strategies: “there’s always some target of something that’s got to be handed, the phonics screening for the Year 1” (Anne, final interview).

5.3.4 Content-Focused Curricula

Time was also short in terms of leaving enough space to follow children’s ideas and initiatives, and some teachers felt they needed tightly controlled lessons so that they could fit in all the learning that needed to happen in order to fulfil curriculum requirements (and, as described above, test-measured standards and data targets). The National Curriculum was described as heavy and “hefty” (Anne, initial interview) and “too much to try and fit in” (Helen, final interview). Teachers felt that their need to fit in all the requirements made it more difficult for them to follow children’s lead: “I find that I keep drawing them back to what I want, to what I want them to do, rather than just letting them” (Julie, final interview). Indeed, the teachers who had given children a lot more space for self-directed learning and free play, found that they were beginning to worry about fulfilling National Curriculum requirements.

However, the curriculum on its own was not always perceived as a problem. Some teachers found there was enough room within it to leave some space for children's interests. For example, Beth explained that although there are always curriculum pressures, the programme had helped her "look at it in a different way" and that she had found "a different way to approach it and tackle it" (final interview). Lisa echoed this, feeling that the way the curriculum was applied was "kind of in our hands at the school at the moment." (Lisa, final interview). In other words, the curriculum was a pressure that became problematic when it was viewed as a list to tick off rather than as something that could be woven and feed into children's interests.

A number of teachers also noted that there was a big difference between Reception and Year 1 in terms of expectations and pedagogy, and that the Early Years ways of working were much better suited to the strategies we introduced. This is perhaps unsurprising considering much of the practices we drew on came from Early Years pedagogies. Anne noted: "There's more of an opportunity to [follow children's interests] within the way the classroom setup works, more naturally" (Anne, final interview).

In particular, continuous provision was seen by some teachers as Early Years practice that would be difficult to implement in Year 1, specifically because of curriculum constraints, both in terms of the amount and the nature of the content to be taught. Some teachers also noted that these strategies would become even more difficult to implement in older classes because these pressures increase as children go up through Primary school.

5.3.5 School Rules and Directives from Senior Leaders

Whilst many of these issues were not in the school's power to change, top-down prescriptions from SLT (the Senior Leadership Team) exacerbated them. For example, in Lisa's school, teachers had to follow rigid schemes of work for Literacy, with scripted, fast-paced lesson plans. This gave very little space for teachers (let alone the children) to make it their own. As a result, Lisa felt that these constraints meant she couldn't try some of the strategies. Continuous provision, in particular, which would require a change in the way the classroom operates, was seen as impossible. Similarly, asking for uniformity in lesson plans constrained teachers, in particular when linked to accountability. Helen described how the Trust (i.e. the private governing body) her school belonged to expected books to look a particular way and for "the evidence in the books looks exactly the same" in all schools, which meant it was "very difficult to go off on a tangent" (Helen, final interview). Often this

meant these teachers' attempts to try strategies to give children more autonomy were restricted to small portions of the day:

Although our boss has been very keen to let us try things out, not for literacy and not for Maths, not for phonics, not for reading. So it's only a very short amount of time in the day and in the afternoons. (Helen, final interview).

However, where teachers had more freedom (e.g. "what we actually teach is a little bit more down to us", Beth, pre-interview) or where new practices only took up a small portion of the timetable, teachers did not find them so limiting.

Some SLT also promoted ways of teaching which discouraged teachers from following children's lead, and which therefore contradicted with the strategies we were encouraging teachers to try. Vicky noted, for example, that her children's free 'choosing time' (i.e. free play) had been criticised by SLT for not being directed enough. Helen also noted:

They don't like that. They like objectives, specific lesson outcomes that they can see on paper that's done at the end. And I always get in trouble for going off on tangents and not sticking to what I'm supposed to do. So for years I've been told, 'don't go off on a tangent'. (Helen, final interview)

Therefore, underlying the pressure from SLT was the idea that teachers would get "in trouble" if they did not teach in ways prescribed by SLT and could not provide evidence that learning objectives had been met. Indeed, the pressure to have a predetermined learning objective was often dictated from above, with children expected to "parrot back in the middle of the lesson what was the learning objective" (Helen, Workshop 4).

As a result, the practices we encouraged were sometimes described as being "risky" (Anne, CI interview) and "stressful" (Beth, pre-interview), because it was not possible to predict where the lesson was going to end up and whether learning objectives would be met:

You don't know what's going to happen. It could go anywhere, it could go nowhere. I think that's the problem, it could go nowhere ... it looks like they haven't done anything all morning or all afternoon, even though they have." (Stephanie, final interview)

Therefore the programme disrupted what had been the accepted model of teaching in many schools. This made it difficult for the teachers to apply the strategies and created tensions.

5.3.6 The Performative, Normative Culture of the Global Education Reform Movement

Underlying the pressure to meet objectives is a view of learning that is not ideologically neutral, but instead has been described as performative, managerial and neoliberal (e.g. Ball, 2003; Biesta, 2010; Mehta, 2015; Sturrock, 2021) and even aligned with right-wing politics (Watson, 2020). The focus of what has sometimes been called GERM (Global Education Reform Movement, Salhberg, 2016) is outputs and measurables, rather than process and growth. Under this ideology, the purpose of school is therefore to produce certain results, rather than self-actualised and happy children. It also seems to go hand in hand with a “banking system” of education (Freire, 1972/ 2000; hooks, 1994, p. 5), where the memorisation of knowledge, often aligned to the ‘knowledge of the powerful’ (Muller & Young, 2019) is seen as the main aim of education.

The teachers in our programme also used the language of productivity and performativity. Speaking of small group work with a teacher, Claire described it as getting “the writing that we need out of them and then they're ready to go off and they rotate” (Workshop 4), representing a piece of writing as a product to be made for external purposes. Helen spoke of her fear that children would use autonomy to be unproductive. In some classrooms, even children’s talk needed to be productive, as outlined in the section on teachers’ control over discussions. Anne spoke of helping children meet standards as the focus of her efforts.

This view of learning as meeting externally imposed targets also contrasted work with play, the latter seen as not productive, or as needing to become productive (“I’m happy with play as long as it’s got a purpose” Helen, final interview). In classrooms where children were allowed some time to play, the play was seen as needing to be on-topic, and children were sometimes diverted from their play to refocus them to the kind of play adults wanted to see. For example, in the continuous provision of Claire’s classroom, children who were not pretending to be at tea with the Tiger Who Came to Tea (a famous children’s story) were stopped and told what to do by a TA (Claire, final lesson 2).

5.3.7 Fighting the Pressure

In a context where teachers are themselves under such pressure and scrutiny, it is hardly surprising that they should want to tightly control children's learning. Indeed, other research has shown that high stakes environments lead to more controlling, teacher-centred teaching (Reeve, 2009; Valli & Chambliss, 2007) and to curriculum narrowing (Berliner, 2011), as well as making it more difficult for schools to innovate (Greany & Waterhouse, 2016). However, as we have emphasised, there were important differences between schools and between classrooms as to how much autonomy teachers were able to give children. Whilst part of this was down to teachers' own beliefs (such as their need for control and their alignment with the banking, target-driven model of education) there were also other factors which helped teachers to resist the pressure and allowed them to give children more autonomy.

Trust and Teacher Autonomy. In contrast to schools where the SLT exacerbated institutional pressures, in some schools senior leaders created protective spaces. Where teachers had large amounts of autonomy, they could experiment more and have flexibility to do what they thought was best. Both Beth and Claire noted that, whilst teachers were still bound by objectives, they had freedom as to "how you approach it and how you go about it" (Beth, final interview).

However, even in schools where SLT created pressures, they also sometimes gave teachers space to experiment because of their involvement in the programme. Helen remarked that her head teacher had been "very supportive" in allowing the teachers to experiment and provide reassurance they did not have to follow all the usual rules imposed by the Trust, for example around uniformity of books. Helen felt this has helped alleviate some of "the additional pressure of worrying about not fitting everything in" (final interview).

Teachers often described their headteachers as supportive if they allowed teachers to stray from the trodden path – that is to say by giving them more space, rather than more structure. The importance of teachers' agency in their practice and professional development has been well-evidenced (e.g. Edwards, 2005; Lipponen & Kumpulainen, 2011; Molla & Nolan, 2020); what the present research suggests is that it is also essential to children's autonomy, because it requires teaching to be adaptive and responsive, and therefore teaching cannot be scripted and uniform.

Coupled with autonomy was trust – the idea that senior leaders trusted teachers to do what was best was key in helping teachers embrace the practices we introduced. This trust was conditional on teachers being able to explain why they were using particular methods, emphasising the teachers’ ability to make judgements as professionals. However, it was also conditional on good results, as Beth emphasised:

Don't get me wrong, she'll still come in and go right, what percentage of children are going to pass their phonics check and how many children have you got working at greater depth and what mastery thing are you doing (Beth, Workshop 3).

Therefore the need to meet targets and perform on tests was still present, even when head teachers were supportive and allowed teachers some flexibility – teachers were allowed to give children more autonomy as long as the results were still there, which was sometimes a source of tension.

The Stepping Stones Programme as Permission to Do Differently. As argued earlier, many teachers spoke of learning as accumulation of knowledge and of the need for productivity in order to meet objectives and standards. However, at the same time, they also appeared to hold views congruent with learning as process and the need for children to be participants rather than recipients of learning – I later discuss the dissonance this created. Therefore, whilst they defended their use of teacher control, many teachers also felt the programme was aligned with their own goals for teaching and it gave them confidence that giving children more autonomy was beneficial:

I think what’s come out that is really positive is that I am, although I’ve always known it, it’s very clear that I have to find a way to give the children much more ownership over what they do and not be so restricted by, this is the end outcome. (Helen, final interview)

Teachers reported the programme had given them confidence (Claire, final interview) and Vicky said it had made her happier, as it allowed her to be “much more of the teacher that I want to be rather than the teacher that I feel I've had to be this past year” (Vicky, final interview).

This increased confidence, as well as the status afforded by the programme itself, gave teachers permission to be subversive. Some teachers saw the programme as an opportunity to experiment in ways they would not usually be permitted to or be used to (“having the opportunity to have management say go ahead and give it a go”, Helen, final

interview; “I don't think I'd have been as experimental.”, Vicky, final interview) or to try practices that were previously frowned upon. Some teachers even began to stand up to others in positions of authority in order to protect the practices they were trialling: “Someone came and asked for a Science plan for what we would do for the rest of the year. And I said you can't have it because I don't know what the kids want to do yet.” (Beth, final interview).

The programme was also a space away from senior leaders, where they could question and even make fun of practices that had been imposed on them. In one of the Community of Practice workshops, the teachers mocked the futility of explicit learning objectives that they could not deviate from, but were judged on during observations, and which children were required to know:

Helen: So what were you told in the first two minutes that you will be learning today. And you will not learn anything else. Because that's the thing you were told to learn. So remember it because next week somebody is going to ask you in the middle of the lesson what your teacher told you and wrote on the board.

Kirsty¹: And [senior leaders] will come in all sweet and smiling. So, what are you learning today?

[teachers all speak at once]

Stephanie: And then they say something random like, Maths, and you're like, No it's Phonics!
[laughter]

Kirsty: Or they walk round to that one child that you know is just going to say snack time.

Unidentified teacher: Or she says we're doing money but they're making a fire engine.
[laughter]

The laughter and effervescence of this short interaction show that this was an experience that the group could all relate to, and the programme gave them space to distance themselves from such practices they disagreed with. It also highlights that teachers' own beliefs about teaching and learning were complex, and whilst they largely wanted to be in control of lessons, they also doubted the usefulness of some of the controlling objective-driven practices they had to use.

¹ Kirsty was one of the teachers who dropped out of the programme because of unforeseen circumstances and not included in the analyses, but she took part in three of the workshops and therefore in the discussions.

5.4 Pockets of Space for Children's Choices and Ideas

Even before the programme began, within this ecosystem of teacher control were spaces for children to make choices, participate in decisions about their learning, and contribute ideas. These took different forms depending on the teacher and their context. Therefore, rather than introducing technical or pedagogical innovations, what the programme really did was support teachers in opening up new spaces and consolidating existing ones. However, these practices were not well connected with each other and remained more or less isolated within the broader system where teachers retained the control over the learning.

Before describing these *spaces*, we should note that in our own Theory of Change, it was instead choice which was considered a key 'ingredient' to children's inner motivation, executive functions and autonomy. However, in previous research investigating children's choices, the focus tends to be on the nature of the choice (e.g. Stefanou et al., 2004). These typologies of choice often end up leaving out the child – it distinguishes what the choice *is*, but not what the choice is *for*. In addition, existing theories do not classify choices in terms of *how much* choice a child has, and therefore do not distinguish how much freedom or control over learning the child has. Finally, choice was often narrowly interpreted by teachers as meaning choice between options, whilst challenge was often misinterpreted as making activities more difficult. Therefore, as the analyses progressed, we had moved away from focussing on choice and challenge per se, and instead focused on opportunities for children to experience autonomy more generally.

Similarly, whilst many of the strategies supported autonomy through choice or challenge as well as directly, mapping teachers' attempts to use the strategies would create difficulties for the present analysis. First, the strategies were not exhaustive and only starting points for teachers to begin to put into practice opportunities for autonomy. Secondly, how those strategies were interpreted and enacted by different teachers varied – as we have seen, what one teacher called elaboration might not have been what others had in mind, including ourselves in the research team. In particular, seemingly similar practices differed in terms of how much autonomy they offered to children because of the way they were enacted and the way they were received by the children.

Therefore, rather than adopting an existing framework or classification around choice, I focus here on the way different practices afforded different kinds and amounts of *space* for the child to take ownership of their learning – which is what we had always meant by 'choice'.

5.4.1 *Small Choices Bounded by Teachers*

A number of teachers gave children limited choices, that is to say choices that were strictly bounded by the teacher. For example, this included option choices between tasks, as in one of Beth's mid-point lessons where children could choose which Maths exercise they did. Whilst children had choice, there was a limited number of options to choose from, and all the exercises focused on the same aspect of the lesson, namely practicing exam-type word problems around arrays. In this sense, these were similar to what Reeve et al. (2003) call 'option choices'

Small choices also included children having a choice of materials for a specific task. For example, in lessons where Beth and Lisa had tasked children with making sock puppets, children could choose the design of their sock puppet, and they could choose what materials they used. Whilst these were more than option choices in that they allowed children to express preferences and ideas through the choices they made, they were nonetheless limited in how meaningful they may have been and were limited to what Stefanou et al. (2004) call procedural choices over of the task.

In some cases, these choices were inconsequential, in the sense that they had little bearing on what the child produced: for example, in one of Danielle's initial lessons, children could choose what colour paper they used to make a card. Such choices may have increased children's sense of control if the child had a strong preference. Nonetheless, this preference could only be there because teachers had given a choice – the children may not have expressed a preference for a particular colour had they not been given the option. However, perhaps the importance of such trivial choices for giving children a sense of power is not to be underestimated. Indeed, research on choices suggests their effect is particularly strong for children and Patall et al. (2008) suggest this may be because children have otherwise limited opportunities to make choices for themselves.

Whilst all these small choices may have mattered to some children, these choices could also be trivial because the children did not care about which option they chose – in other words, the *outcome* of the choice did not matter. Whilst it may be difficult to know in advance whether a choice will matter to a child, the decontextualised nature of option choices – children had little ground for choosing one over another – meant they lent themselves more to being trivial.

Overall, what characterised these limited choices was that they did not allow children to take charge of the learning and therefore they were somewhat tokenistic. They might have

increased interest if they allowed a preference to be expressed, but the teacher remained in control of what the purpose and outcome of the learning was.

5.4.2 Space for Self-Direction

In stark contrast with trivial choices, in some classrooms children had opportunities for self-direction, in particular through continuous provision and through open-ended tasks.

Continuous Provision and Play. Continuous provision has not been formerly defined (Trollope, 2019), but in our study teachers used it to refer to situations where children had a free choice of activities during a long (at least 45 minutes) uninterrupted period of time. This included opportunities for free play as well as more structured activities that were set up on the day, which may go against what others see as ‘true’ continuous provision (e.g. Early Excellence, 2021). For example, in Anne’s Reception/Year 1 classroom, children could play outside in a playhouse or with toys, but in every session there were a number of activities linked to the topic set out at different tables. In her lesson on minibeasts, most activities were minibeast-themed, for example spinning wool around table legs to make a web and balance a spider.

In older classrooms, often there was an expectation that children’s play would link to the day’s learning objective so it was not entirely unrestricted. In one of Julie’s lessons (mid-point lesson 2), she introduced what were the options for learning in the afternoon, which were all Maths-based and related to the day’s learning objective (halving). This included using drawing shapes to show how many ways they could halve and making pizzas with each half having different toppings. Children could also continue observational drawings from the morning, but were encouraged to look for halves and draw halves in what they saw. However, children also had access to more unstructured play alongside those activities. Whilst children were accessing the provision, teachers were usually leading a structured session. For example, during continuous provision after her lesson on minibeasts, Anne worked with a group of Year 1 children on “asking questions about spiders” and finding out the answers on computers. Julie and Claire also worked with a small group of children on a teacher-set task, for example a worksheet or a piece of writing.

When these activities were more limited in scope – for example when children could only choose between a set number of activities – teachers sometimes called this a carousel. This was the case in Danielle’s science lesson on leaves (mid-point lesson 2), where she had

set up a range of activities and investigations at different tables all encouraging children to observe and represent leaves in different ways: pencil drawing based on observations through microscopes and magnifying glasses, cutting card and gluing on straws to match the veining of leaves, drawing using pastels or felt tip, painting using powdered paints the children mixed themselves, and engraving polystyrene to make leaf prints. The children could visit any table they liked during the 45-minute period of the lesson.

However, despite not being entirely unbounded, continuous provision and carousels allowed children a great degree of freedom and opportunities for self-directed learning. For example, in Claire's initial lesson, some children were sitting at a table, writing independently, unprompted. In her final literacy lesson, based around the story of the Tiger Who Came to Tea, a child had chosen to draw a tiger, whilst others had chosen to make a story board using the book as a guide. In one of Anne's initial lessons, a child decided to line up toy snakes the teacher had used during the input (about measuring), and to line them up to compare the sizes.

It also allowed them to engage in social play and for spontaneous cooperation. For example, in one of Anne's initial lessons, the children had made themselves an obstacle course using stepping stones and the playhouse, taking it in turns to go through the course and negotiating the rules, sometimes with the teacher's help. Another example of spontaneous cooperation in self-directed activity is the number line children tried to create in Julie's class (described in more detail in her story of change), where children decided to create a number line and worked together to find out what numbers were missing. The freedom of continuous provision also allowed children to engage in spontaneous discussions. For example, on one occasion we noted that children were naming plants and discussing what could have eaten the strawberries as they watered the plants (Claire, final lesson 2).

Open-Ended Tasks. Whilst continuous provision allowed children to choose what to do, open-ended tasks allowed children to choose where to take it. Open-ended tasks tend to be conceptualised as tasks where there is no single way of solving a problem or where there may not be a single solution. These tasks therefore involve problem solving and higher order thinking (Osana et al., 2006) and have been described as requiring teachers to relinquish control over aspects of the task (Chan & Clarke, 2017). Here, I do not refer to tasks where children could choose how to do the task in terms of materials or method (as in making puppets out of socks), but where the end-point of the task or the strategies for reaching it were not pre-determined; therefore where children could decide what direction they took the

learning in within the task or what the end result was. For example, in one of Beth's lesson, the children had to build the tallest tower out of newspaper. The task was open in terms of what children built, what strategies they used and what shape the tower would be. Similarly, in one of Helen's lesson, the children were building a beanstalk (which they had wanted to have), and Helen gave them a period of time to try some ideas for making the beanstalk. Whilst the end result was not what the teacher had wanted, it gave the children a chance to experiment and to direct their own activity.

Typically, open-ended tasks involve exploration and investigation (Viseu & Oliveira, 2012), but this was rarely the case in our programme. Overwhelmingly, when teachers set their class what they saw as open-ended tasks, they were centred around a particular product to create, rather than being truly open in terms of where children could go with the activity. Viseu and Oliveira (2012) also argue that in order to be effective, open-ended tasks need to be guided and followed up by discussions, which we rarely observed. One exception was in Helen's beanstalk activity, where the teacher asked a number of questions to help children think through how they might build their beanstalks, e.g. "What would we need to do with these to make them into a beanstalk?"; "Will that stop them from falling apart?"; "What could I do to help you reach the top" (Helen, mid-point lesson 1).

Space to Explore. We observed a number of lessons where the task was not only open-ended, it consisted mostly of free exploration. In those lessons, some teachers described their aims as wanting to see what "came up" (e.g. Julie, mid-point lesson 2). In particular, four of the teachers described how they had given their children a chance to explore the school grounds as part of their topic on plants or on minibeasts. Sometimes, the exploration was guided by the requirement or invitation to record what they had seen, for example by giving children clipboards or cameras.

These periods of exploration stimulated rich discussions, linking children's thinking to the learning they were doing in class. For example, Helen encouraged children to think about why animals might like to live in the place children had found them, whilst Beth asked children about how they knew what certain flowers were or what parts of plants they had observed. These periods of exploration were also opportunities for teachers to see what children were interested in: Helen made a note of children's questions and Danielle explained, "I wanted the ideas to come from them. I wanted to find out what they're interested in" (Danielle, CI interview).

Some teachers also saw continuous provision activities as an opportunity for children to explore different concepts, by applying them in diverse ways. Space to explore also involved children being allowed to experiment, and to make and learn from mistakes. For example, Lisa described a child who wanted to make a sock puppet out of tin foil. The teacher thought this was not going to work and probed the child to think about the material. When the child maintained he wanted to use tin foil, Lisa allowed him to try it anyway, and to learn first-hand about the properties of the material, even though she thought the child was using the wrong material. Similarly, Julie and Beth allowed children to experiment watering or growing plants in various products unlikely to work (e.g. marshmallow, coca cola) so that they could experiment, and reflect on why the results were not what they had expected. Experimentation driven by intrinsic motivation is considered an important mechanism for children's learning (e.g. Bonawitz et al., 2011; Gopnik, 2012; Taffoni et al., 2014), and in particular surprise and anticipation have been shown to be an important aspect of this kind of learning (e.g. Andersen et al., 2021; Kuhn & Ho, 1980; Stahl & Feigenson, 2015). This suggests that by allowing children to experiment and test their ideas out, and by guiding and supporting these attempts, teachers were facilitating an important way in which children learn about the world.

5.4.3 Following Children's Lead: Space for Children's Input

Teachers also provided children with opportunities to have autonomy in their learning by following their lead. This required loosening the frame of control around them and being responsive to children's ideas as they emerged.

Going Off-Script. Sometimes, this happened in-the-moment, in the sense that a child's comment, question or initiative was attended to and led the learning in a direction the teacher had not planned in advance. For example, in her guided reading sessions with small groups of children, Danielle often left room for conversations that were not entirely on topic. Sometimes these conversations were more social in nature, but sometimes they allowed the teacher to make links with more formal elements of their learning. For example, the children shared stories from their home lives which related to a number of important topics in the curriculum, such as safety and habitats for animals, and the teacher used these stories to probe children's knowledge and prompt their thinking (Danielle, final lesson 1). Similarly, despite her worries about going on "tangents", Helen also sometimes pursued children's ideas

and comments even if they were unrelated to the learning at hand. For example, in one lesson on positional language, she drew a square as a frame, and a child commented that it was a rectangle rather than a square. Helen then asked the child to elaborate on the difference between the shapes even though this was not the focus of her lesson, encouraged the child to be more specific with her answer and then rephrased what they had said (Helen, initial lesson 1).

At other times, this meant teachers let go of their preconceived plan to let the children continue with something they were engaged in. For example, Claire described how some children had begun acting out *The Tiger Who Came to Tea* in the continuous provision after the class had read it and were “engrossed” in their role play, with one child dressed up as the tiger. The teacher was working on writing dialogues with groups of children, but decided to let those children continue with their play rather than interrupt it: “I thought actually, they’re doing the dialogue because they’re acting out the story, so why stop them to come and do the writing? I can catch them to come and do the writing tomorrow.” (Claire, final interview). Therefore following the children’s lead and being responsive to what interested them in the moment required teachers to go off-script.

Flexible Planning. In other cases, teachers had planned to draw on children’s ideas and interests to plan future lessons, something they called ‘flexible planning’ (in reference to the strategy we had introduced). In this case, children’s ideas were not so much followed in the moment, but noted down and used as the basis for future lessons. This occurred, for example, when teachers introduced a dead plant as a hook or took children to a garden to stimulate questions, and then used those questions as the basis for further investigations.

Overall, teachers followed children’s lead by being responsive to what had engaged them and being flexible in their teaching to accommodate it. It meant children’s interests were taken into account when making decisions about what to learn next. For example, Danielle described that she continued with the same topic because the children were so interested in her ‘living things’ topic.

However, whilst the teachers were responsive to the children’s interests and what they wanted to learn, it was unclear how explicit this process was to the children and whether they experienced it as empowering. For example, when Beth explained how she had used the children’s questions to plan her topic on plants, she never explained whether the children were aware that she was trialling this new practice and was intentionally drawing on their ideas for planning. However, she indicated that the children were more interested because the

5.4.4 Spaces for Dialogue and Meaning Making

Teachers also opened up spaces for children's ideas by opening up the conversation. The strategy of 'elaboration', which was one of the first introduced to teachers, was explicitly trialled by four teachers and others also reported they felt it was part of their existing practice. Although the focus was on deepening interactions and helping children justify their answers, it also served to make space for children's ideas.

However, elaboration in itself did not always lead to more space for children's ideas, in particular when teachers were still focused on a correct answer. For example, in Lisa's attempt at elaboration (see her story of change), she steered the children towards the conclusion she wanted them to reach. Because her main activity hinged on this conclusion, she had a script of what she wanted the children to say and did not engage in children's own ideas. Sometimes, teachers also saw elaboration not as a way of giving space for children's thinking, but as a strategy for forcing children to practice oral skills: "They tend to give you just one-word answers, or a couple of words, so you're constantly trying to get them to speak to you in sentences, or, to give you a full answer." (Stephanie, final interview).

Therefore, it seemed to be that it was when teachers engaged children in *dialogue* (Mercer et al., 2009; Vrikki et al., 2019) and *sustained shared thinking* (Siraj & Asani, 2014; Siraj-Blatchford et al., 2008), building on the children's ideas and helping them make sense of their thoughts, that teachers were successful in making space for children's ideas. For example, in one lesson, Anne had brought a plant she had dug up, its root intact, and had engaged the children in a conversation about the parts of the plant. Whilst the back-and-forth of questions always went back to the teacher (rather than children discussing between themselves), Anne allowed space for children to notice and provide ideas, for example about the function of the leaf or why there was mud on the roots, without steering them, even though there was specific information she wanted them to learn. However, as argued earlier, we rarely observed this and more frequently conversations were steered towards a set of correct answers the teacher wanted to reach.

Another way in which teachers opened up children's discussions and meaning making was by inviting different interpretations and ways of seeing. For example, Helen reported a lesson where she had showed the children 'The Snail' by the artist Matisse, but decided not to tell them what it was called and instead asked them what they thought they could see. The artwork therefore acted as an open prompt for meaning making and discussion rather than a model that could be replicated. In her debrief, she described in detail the variety of answers

she had elicited and the rich discussion that ensued, as well as the creativity in the artwork the children had produced, which she believed would not have occurred if she had focused on the correct interpretation of the artwork and told them it was called *The Snail*:

Every one of them would have cut the paper up, stuck it in a spiral, and every one of them would have looked like snails. And then, there's nothing unique about that, they haven't interpreted it in their own way ... But what was really lovely was that everybody's were their own. (Helen, mid-point lesson 2 debrief)

5.4.5 Changing Practice by Opening Spaces

We have noted before that teachers did not make large changes to their practice. Many of the strategies we introduced to teachers were familiar to them. Indeed, a number of them were described as natural to teachers, or linked to practices they already used, like questioning. Therefore, for many teachers it was not a question of overhauling their practice – which in any case was not our intention – but of opening new spaces or consolidating existing ones. Beth commented that the programme had made her more open to children's ideas and follow their lead more spontaneously than she had before. Many teachers noted the programme had helped them reflect on their practice and had brought their attention and awareness to aspects of their practice they had not always considered. For example, Lisa spoke of being more mindful of the practices.

However, on the whole these spaces remained disconnected. Whilst Julie and Claire had changed the organisation of the classroom to allow for some free choice of activities (by far the largest change any teacher had made), it did not go hand-in-hand with a questioning of who was in control of learning, or an increase in children's autonomy in other areas. Many of the strategies were instead tried in a punctual fashion or restricted to one afternoon a week in topic or science, whilst the rest of the time teachers relied on teacher control. It resulted in a piecemeal approach which made it even more difficult to maintain the practices in the face of tensions, and likely exacerbated these tensions. It is those tensions I now address.

5.5 Tensions Between Ways of Working

As described above, teachers' attempts to give children more autonomy were not marginal, but often firmly ensconced into everyday practice. However, because these practices were disconnected and at odds with the ecosystem of teacher control, they did not

mesh well with the rest of the teachers' practice and at the seams between the two were areas of tensions and contradictions.

5.5.1 Operating With (and Within) Contradictions

The teacher-controlled classroom is, by definition, at odds with children having space for autonomy in their learning. At times, this created situations where teachers sent children contradictory messages or held contradictory expectations. For example, teachers both emphasised that children should come to them for help, and that they should be independent. In one literacy lesson, Beth told children who had brought their books for her to review their work: "I don't want people now, we talked about this, you are now responsible for checking your own work" (Beth, final lesson 1). However, a few minutes later she accepted children who had brought their books to her and gave them feedback. Therefore, whilst at times she wanted independence, the message she sent the children was still that she was in charge, and therefore this contradicted her requests that they act independently.

Teachers also wanted children to show creativity and self-reliance in solving problems, but only at times when it matched the teachers' own goals, and when it led children to the desired objective. Even before the programme started, Helen worried about children's lack of independence and problem solving, but in the lesson where children were making a beanstalk, she described the children's attempts in negative terms because they were not what she had wanted out of the task. Her aim had been for children to work collaboratively to make a large class beanstalk, but instead the children cut up leaves on their own or made their own small beanstalks. She concluded "So they will need an adult to work with them this afternoon and teach them how to cut it, what sizes they're going to need, how they're going to construct it, the fact that they're going to need two of them to hold it together while somebody else is taping it" (Helen, mid-point lesson 1 debrief). Therefore, whilst acknowledging that the children had, in fact, worked without help on making parts of a beanstalk, she decided they would need to be told what to do, "a best possible way to do it", because they had not made the beanstalk in the way she had intended. She also did not consider the impact this might have on children – to have a chance to create something, only to be told they would have to start again because there was a better, teacher-approved way to do it. Therefore, teachers wanted children to rely less on the teacher and solve problems out of their own initiative, but only as long as they were compliant with the teachers' wishes, in a context with a large focus on correct answers and methods.

These contradictions seemed to be the expression of dissonance between different beliefs teachers held and different goals they pursued simultaneously. Teachers were often ambiguous and contradictory in their positions regarding children's autonomy. Teachers emphasised the need to be in control and to have learning objectives but at the same time they criticised the pressure and accountability around these learning objectives. Most teachers were enthusiastic about the goals of the programme, but nonetheless acted in controlling ways. Helen, in particular, struggled with the tension between these opposing forces. At times, she was very clear on the need for teachers to have control, to have clear outcomes and for children to challenge themselves rather than "just play". But at other times she was very positive about the practices we encouraged, despite her struggle:

When I first heard it, I thought 'oh no, this is not going to be for me', because, I don't get along so well with that. Again, with my whole control of, this is the outcome at the end, but actually I can see that it really benefits the children. (Helen, final interview)

As argued above, part of that ambiguity was fed by the pressure from above, which steered teachers towards controlling practices, and partly depended on teachers' own underlying beliefs. For example, despite at times acknowledging that it "should be" how she should teach, Anne seemed to experience the programme with a larger amount of cognitive dissonance compared to other teachers, as the programme conflicted with her existing practice and beliefs about teaching. She described feeling "at sea" and "all over the place" because of the conflict between the practices we were encouraging and what she had "always done, which has worked" (CI interview). As Anne highlighted, this dissonance sometimes resulted in uncomfortable feelings: "I think I felt really unsure about it and didn't like it" (Anne, final interview). Helen also mentioned that these tensions brought up difficult feelings and her struggle in navigating those contradictions in practice despite "knowing that it's the right thing, and it's always been the right thing" (Helen, final interview).

5.5.2 The Case of Children's Lack of Independence: Transitioning Towards More Autonomy?

A particular problem that teachers faced when giving children more autonomy was that they felt children sometimes lacked independence. I use the word independence here purposefully as distinct from autonomy, because independence implies acting on their own and self-reliance, rather than self-direction or self-determination. Children working on a

teacher-set closed task without adult help are working independently, but they have little autonomy.

A Lack of Skill. Teachers often interpreted children's struggle with independence as a lack of skill – children did not know how to work on their own – and to remedy this issue, many teachers discussed the need to teach children to be independent by teaching specific skills (though teachers were not explicit about what those skills were). This also framed the issue of children's independence as one of transition: as teachers changed their practice, they needed to transfer the responsibility for learning or for doing onto children, and therefore autonomy required children to have some amount of independence. Indeed, we observed this implicit transfer on many occasions. Often, this was through questions which redirected the onus onto the child:

One child calls out, 'I don't have any space on my whiteboard' and the teacher asks, 'what do you think you could do?' She repeats twice until the child cleans her board. (Stephanie, mid-point lesson 1)

At other times, teachers made suggestions for what children could do; for example Claire suggested a child talk to his partner about an issue, rather than ask the teacher to solve it for him. (Claire, final lesson 2). Stephanie also spoke of trying to increase children's responsibility for getting resources they needed by leaving options open for them, for example by not putting the playdough out because children could get it themselves if they wanted it (Stephanie, final interview).

This is in line with other research that suggests that children can learn to self-regulate in classrooms where there are opportunities for children to have autonomy and where teachers slowly build up children's capacity to self-regulate (e.g. Perry, 1998; Perry, 2013). In addition, these teachers' conception of autonomy relying on independence as a skill relates to research on metacognitive and executive function skills and self-regulation strategies (e.g. Diamond & Ling, 2016; Fuhs et al. 2013; Robson, 2016; Torres et al., 2018) and suggests perhaps teachers would have appreciated to engage in this body of research.

However, these nudges towards greater responsibility occurred in an otherwise teacher-controlled context, which (as argued above) sent children contradictory messages. Teachers rarely considered how the broader environment supported or hindered children's independence. The culture of the classroom around independence was often ambiguous, as highlighted above, as teachers wanted to retain the control *as well as* for children to be

independent. For example, in the example of the child unsure what to do with their whiteboard in Stephanie's class, it may have been that the child was unsure whether they were allowed to erase their previous work. In classrooms where children are reprimanded for showing initiative, it may be difficult for them to ask for resources that have not been put out for them, as Stephanie suggested they could. In addition, it is possible that children had been socialised into a passive role, and therefore when teachers gave children autonomy in a context where the teacher was otherwise in control, children still maintained their role as passive and helpless despite the opportunity to become active.

Other pedagogies may provide clues as to how this could be different. For example, in Montessori pedagogy, the environment (including the teacher) is seen as key in supporting children's ability to choose tasks and self-regulate. Children's failure to cope with self-governance is considered to often be an external failure (a lack of support or the wrong conditions) rather than an internal one due to the child's incapacity to have autonomy (Frierson, 2016). In other words, all children, whatever their level of maturation, can experience autonomy given the right conditions. Similarly, some teachers questioned whether the classroom culture might be responsible for children's lack of independence and that the children had got used to teachers telling them what to do.

In addition, the conflation of autonomy with independence meant that teachers did not consider ways in which they could support the experience of autonomy of children who were still highly dependent on adults or peers – in other words, the possibility of a high autonomy/high dependence situation.

A Need for Structure. Teachers also suggested that children's struggle with independence was a sign that they had been given too much freedom, which they "could not cope" with. Often, this explanation was associated with a concern for disruptive behaviour. For example, Anne and Lisa described lessons as chaotic and out of control because they had tried giving children more freedom. This often led teachers to argue for a return to more teacher control. This is because teachers explained children's struggle as a *need* for structure and a clear framework and therefore the solution to this issue was to increase 'structure'. Structure has indeed been found to support learning, especially in autonomy supportive environments (Hospel & Galand, 2016; Jang et al. 2010). However, structure may refer to a variety of practices, and often teachers referred to structure to mean teacher control. For example, after her lesson on fair tests, Anne concluded she should have given the children fewer choices and been more directive in how they should do their test, instead of

considering children's ownership over the inquiries and their interest in the questions they were answering. By contrast, Beth supported children's inquiries on plants by facilitating them (e.g. getting the resources, encouraging discussions), or by proposing experiments to test out their ideas. Structure that is supportive of autonomy might also take the form of agreed rules for talk (Warwick et al., 2013), a highly ordered environment (Lillard, 2019), tools and procedures (Bodrova & Leong, 2018; Lillard, 2019) or scaffolding (Perry, 1998; Rojas-Drummond et al., 2013).

Further Considerations for Children's Lack of Independence. Whilst the above explanations may partly explain children's struggles when given more autonomy, I argue they are not sufficient to fully understand them. If it was only a lack of skill and/or a need for structure that caused children to struggle with independence, we would observe two things: first, as children got older, they would be given more freedom (as they would be more able to cope with it); and secondly in settings where children have much more freedom, they would struggle more. I argue this is not the case.

Whilst older children are more able to self-regulate (Best & Miller, 2010), clearly very young children in play-based, continuous provision settings have much more freedom than older children in highly teacher-led classrooms. In addition, in settings where children have (by the standards of schools) extraordinary amounts of freedom, children do not struggle with independence. For example, in self-directed learning or home settings, children frequently learn about diverse and complex areas independently without the need for highly controlled lessons (Thomas & Pattison, 2013; Traxler, 2015).

Finally, some children in the classrooms we visited clearly struggled with independence even in tightly-controlled situations. Indeed, Helen and Lisa's concern for children's lack of independence pre-dated the Stepping Stones programme. I observed that teachers were often busy supporting children to help them get tasks done – and as the only adult in the room, they were often spread thinly. They often had long lines of children waiting for their help or for feedback. They also sometimes worked one-to-one with children, giving step-by-step instructions, where children were entirely reliant on the teacher for completing the task:

She gives prompts to the child, e.g. "I am..." and the child finishes sentence verbally. The teacher points and says "I", and the child writes. The teacher asks, "What's your next word going to be?" and helps the child word by word, and reminds her of finger spaces by pointing. (Beth, final lesson 1)

In those situations, children were heavily reliant on the teacher to complete the task or to acquire information they needed. Therefore, I propose children were not independent because *they did not know what to do*, whether or not they had autonomy. I discuss the implications of this, and what teachers did that helped solve this issue, in section 5.6.

5.5.3 When Children Don't Choose What the Teachers Want

The learning objective mindset also meant that when teachers gave children more freedom, they nonetheless had expectations for how the children should use that freedom, and in particular that it should be productive. When they did not, teachers felt this was a negative consequence of having too much autonomy. For example, when the Reception children did not choose her tasks of making minibeasts in continuous provision, Anne concluded she should have made it a compulsory task. Helen mentioned she often steered children towards specific activities she wanted children to do (Helen, final lesson 2 debrief).

Play in particular was often viewed through the lens of learning objectives (when it was at all present), and I have previously described how children in Claire's class were redirected in their play so it would match the topic of the Tiger Who Came to Tea, and how Helen wanted play to be purposeful. Vicky also mentioned that her head teacher was pressuring her to stop free play and instead prescribe focused lesson-related (but playful) activities such as making "cakes out of LEGO that are split in half" following a lesson on fractions (Vicky, initial interview). There was therefore a sense that play that was not directed at explicit learning objectives was frivolous and undesirable. Those teachers who did allow time for play during classroom time therefore often restricted it so it would be productive. For example, Helen explained adults needed to intervene to stop children playing unproductively ("the car doesn't need to go up and down the walls", Helen, final lesson 2 debrief).

This highlights the tension that arose between different goals for children's autonomy and children's play. In opposition to a view of play as the "naturalistic and free activity of childhood" (Wood & Hedges, 2016, p. 390) is an adult-centric instrumentalised conception of purposeful play (Bautista et al., 2019), which can lead to practices that constrain, narrow or divert children's play (McInnes, 2019; Wood, 2014).

Associated with the idea that play needed to be productive was the idea that play was opposed to challenge, that when children "just" played, they did not challenge themselves

and needed to be guided to do something else. Because challenge was one of the core ideas of the Stepping Stones programme, it was one teachers often mentioned, but the way they spoke of it echoed the notion of high expectations some teachers referred to (e.g. Lisa). In particular, teachers understood challenge as children choosing to do academic work, difficult work, or novel work. In particular, some teachers commented that when children chose tasks they were familiar with, or which played to their strengths, this showed a lack of challenge. For example, Julie reported that children who had presented their work as posters because they were good at drawing (leaving other children to find the information) were demonstrating a lack of challenge by choosing something they were “comfortable with” (Julie, final interview). The teachers’ descriptions of challenge were centred on the idea of children going over and beyond, and teachers often spoke of pushing children, or wanting children to push themselves: “They needed a lot of pushing from me to say, no, it needs to be something challenging” (Stephanie, final interview). Therefore, some teachers were concerned that if children had more freedom they would not challenge themselves (“Some children just want to do the easiest option it's where to eke in that challenge all the time,” Lisa, final interview).

Some teachers also reported the opposite problem – that when given the choice, some children chose tasks that were too difficult for them. Some teachers suggested that it could be because these children did not want to be seen to be doing easy work – this was potentially reinforced by teachers implicitly or explicitly telling children that doing difficult work was desirable (such as teachers emphasising the importance of challenge or rewarding children who had challenged themselves). Therefore, it seems that some children may have struggled to appraise the difficulty of tasks, or that the reason they chose a particular level of difficulty was to do with reasons unrelated to the learning itself.

By contrast, some teachers noted that moving away from teacher-directed tasks had allowed children to move away from this mindset, because it did not single out children who did easier work as clearly as differentiated worksheets or ability grouping (“She would just sit there and not do anything ... but now, there’s always something she can be doing,” Julie, initial interview). Similarly, some teachers found that when children had large amounts of choice, such as in continuous provision, they were better able to access learning at the right level for them.

However, teachers often implied that the reason children chose easy tasks was motivational: they often did not want to do work that was hard or new to them. If this impression is correct, it may suggest that when given the choice, children do not always want

to do new or difficult work. Pedagogies where children have a large amount of freedom (such as Montessori) show that children sometimes spend considerable periods of time working with familiar tasks or returning to them time after time (Lillard, 2017). There may be benefits to this increased exposure, such as improved understanding (Crawley, 1999). In other words, whilst challenge is important, children cannot be in a state of challenge all of the time, and giving them freedom may mean accepting that, some of the time, children will not challenge themselves. In addition, evaluating children's freely chosen activities through an adult-centric view may also miss important aspects of these activities because they do not fit with adults' preconceptions of what challenge looks like. Maria Montessori described children's repetition of tasks as stemming from a desire to achieve perfection (Lillard, 2017), which in fact suggests challenge.

Finally, other research also suggests that when individuals are interested in tasks, they are more likely to choose or persist in difficult tasks (Fulmer et al., 2015; Inoue, 2007), whilst a study in preschool classrooms found that children in play-based classrooms were more likely to choose more challenging maths tasks as well as voluntarily engage in academic tasks than children in didactic classrooms (Stipek et al., 1995). It may have been that children in the classrooms we visited did not challenge themselves because they were not intrinsically motivated. This highlights how important it is for teachers to foster children's inner motivation, both in the classroom structure and in the tasks, if adults want children to choose to challenge themselves.

5.5.4 Meeting Everyone's Needs and Wants

Different Students, Different Needs. Even before the programme started, teachers spoke of the difficulty of meeting all of their students' needs when they so widely differed. In classrooms where all children were required to do the same task at the same time, this sometimes caused issues. For example, different children might have different attention spans, or be interested in a task for shorter amount of time than others. Helen mentioned a four-year-old boy who easily became bored with activities ("Ten minutes and he's kind of done. He's found some minibeasts and he doesn't want to do anything else now", mid-point lesson 2 debrief). The programme made this issue more present at times, as in the examples above, perhaps because children's interest in their tasks had become such a focus. In addition, greater autonomy in unstructured activities like exploring perhaps made individual

differences more obvious, whereas in carefully differentiated tasks, the teacher may have been able to cater to different needs.

However, at other times, greater autonomy in fact made it easier for teachers to give children the right amount of support. In particular, in continuous provision children were more self-directed and had more opportunities to play, which they did independently of the teacher or with TA support. This meant the teachers were free to focus on a smaller number of children, and this gave them opportunities to better support, challenge and assess children (“It’s much easier than having the class of 30 sat down” Claire, Workshop 4).

Different Interests. Another issue that arose was that when children were given opportunities to make decisions about their learning, this was often done at the level of the classroom rather than individual children being able to choose something for themselves. For example, when doing enquiries based on children’s questions, only a few questions were chosen. We also anecdotally observed large differences in children’s interest in the same task – for example, in one lesson some children were engrossed in making sock puppets whilst one child was going around the classroom with an undecorated sock, using it as a puppet to say hello to classmates. Similarly, in whole group discussions, not all children were interested, and this meant the teacher was torn between following some children’s ideas and questions and losing other children’s interest. Stephanie also spoke of a child who was very interested in horses and explained that the adults often used this interest (for example in interventions) in order to make tasks interesting for her. However, this only worked when working one-to-one with her, as other children in the class did not share this passion.

However, given that classrooms were made of up to 30 children, all with presumably different individual preferences, I found it surprising that this was not a greater issue. In particular in classrooms with whole-class projects (such as enquiries), teachers seemed to speak of children’s interest and engagement as though the whole class had been on board with the ideas (e.g. “They are all buzzing about it”, Claire, CI interview; “The children were just mesmerised”, Danielle, final interview; “They were gobsmacked that it grew in the dark”, Beth, final interview).

This may be because these teachers’ own enthusiasm masked individual differences which could not be fully explored in the short timeframe of the interviews – that the teachers’ use of ‘they’ to refer to the whole class in fact referred to a majority of children only. It may also be that some features of these lessons were broadly attractive to all children, including to children who are usually less engaged, as the lessons often involved exploring, experimenting

or making. For example, Helen noted that one boy who was usually reluctant to participate had become highly engaged because the kind of activity they were doing (outside, hands-on) was more suited to his interests and personality. Indeed, research suggests there are “triggers of interest” (Renninger et al., 2018), which are more likely to engage children. It may also have been that by being part of a group project, where their voices mattered and their participation was valued, children felt ownership of the project even when their own idea had not been picked (Hofmann & Rainio, 2007).

Finally, it may also have been that the novelty of these practices fostered children’s interest, above and beyond the features of the practices themselves, and that if the teachers had sustained the practices in the long term the novelty – and children’s enthusiasm – would have worn off.

‘You Can’t Always Get What You Want’. Teachers also mentioned that sometimes they could not give children what they wanted, and they worried about the impact this had on their interest and their feelings whilst at the same time reporting that this was something children needed to learn to cope with. In other words, because teachers had opened up the space for children asking for what they wanted, this sometimes led children to make requests that teachers could not meet. Often, this was related to materials not being immediately available and “having the stuff there at the time when they want it” (Beth, Workshop3), in particular as following children’s lead meant teachers did not always know in advance where the learning might go. In particular, some teachers worried that children’s interest would wane if they could not do what they had asked for (“I couldn’t have spent three weeks getting everything ready. It had to be the next week, or their interest would have gone” (Julie, CI interview).

However, whilst this was presented as an issue, some teachers noted that it was impossible to anticipate everything the children might ask for and therefore that it was all right for them to say no or ask children to wait. Some teachers even turned this issue into a learning opportunity: if resources were not available, they involved children in discussions or plans for what to get and how. For example, Stephanie mentioned that her children wanted to have a pretend-play shop in the classroom, so she worked with them to create a list of what they would want to have in their shop.

5.5.5 Resources and Organisation not Fit-For-Purpose

Whilst a lack of physical resources (such as teaching materials or equipment) was sometimes an opportunity, it was also a main obstacle to teachers being able to change their practice towards more autonomy. A lack of resources sometimes meant teachers found it difficult to implement some practices – continuous provision, in particular, required materials for children to access learning through play. Claire described the lack of resources as their “biggest thing” and “a huge barrier” (final interview). Similarly, when teachers wanted children to access information independently, they found they had few appropriate resources and therefore had to spend time curating content. For example, Helen was reluctant to give children iPads to research information about food chains in case they came across inappropriate content (“pythons eating cows” Helen, final lesson 2 debrief), and therefore did the information research for them.

Lack of funding was seen as the reason why classrooms lacked these resources. As a result, some teachers bought resources with their own money (e.g. Vicky, final interview), though some teachers also thought their schools were well-resourced (e.g. Vicky, final interview; Danielle, final interview). Some have argued that child-centred practice is more resource-demanding (Power et al., 2019), though such claims tend to consider resources in a broader sense than the physical resources teachers referred to in this research. In some pedagogies, the importance given to the environment as the ‘third teacher’ means that in these classrooms there is an emphasis on the quality of the provision, which includes but is not limited to physical resources and materials (Lillard, 2019; McNally & Slutsky, 2017). However, in many cases teachers did not only complain of not having enough resources, but also of needing *different* resources from the ones they usually had, such as resources appropriate for children to access independently. The issue was therefore caused by the shift in practice, rather than by the practices themselves.

Nonetheless, lack of adequate resources encompassed more than materials. In particular, the number of adults (a human resource) was a factor mentioned by almost all teachers – either as an affordance because they had enough support, or as an obstacle because they did not have any TAs. This was not only to support children’s learning, but also to help teachers when they needed to deal with challenging behaviour. In particular, some teachers noted that in classrooms with children with Special Educational Needs and Disabilities (SEND), not having enough adults to support them was an important issue:

Two of them run off round the school. And if you're in here on your own you can't be giving children choices and having a fully continuous provision if you've got that happening, it's quite tricky to do. (Danielle, final interview)

However, Beth noted that whilst being the only adult in the room made some practices more difficult (and continuous provision impossible), it still was possible to try some of the practices: “We've just learned to get on with it and we muddle through the best we can” (Beth, final interview).

Lack of space was also an issue mentioned by some teachers, in particular for continuous provision. Those teachers who used continuous provision also noted that on days they had less space (for example, if the weather did not allow them to use the outside area), it made conditions difficult: “We'd got more children in a space, fighting for their own room, and actually, the whole atmosphere changed” (Julie, final interview).

Finally, because teachers gave children more choices, this sometimes meant that they had anticipated a number of possible choices children would make, or they felt they had to be prepared for all eventualities. For example, Stephanie talked about her thought process trying to think, in advance of the lesson, of at least different three resources, including puppets and soap carving, which children might want to access so she could get them ready. This made such lessons particularly resource- and time-intensive (“It's time consuming, you can spend the whole lunchtime just getting stuff ready for one lesson.” Beth, Workshop 3). For other teachers, when they wanted to have a lesson where children had more autonomy, often this required a different set-up to their existing classroom arrangements, and teachers often spent considerable amounts of time setting up. For example, Danielle noted that for her carousel lesson she had arranged to have the lesson at the very start of the day so that she could have time to set up all the necessary resources. This impacted on teachers' already heavy workload. However, Claire noted that by using continuous provision, the increased time spent on getting resources ready was balanced by less time spent planning.

In addition, sometimes it was not only the resources, but the organisation and structures of classrooms that hindered teachers' attempts to give children more autonomy. Strict timetables imposed by schools meant teachers lacked the flexibility they needed to follow children's interests. Similarly, often classrooms did not have structures and routines that empowered children. Changing to continuous provision had not only required different resources and a different timetable, but also an entirely new way of organising classrooms through the creation of different areas, different roles for the TAs, and the implementation of

a rota system the classes used. Such changes had taken both the teachers and the children some time to adapt to and Claire felt it was still “an ongoing thing” (final interview).

Therefore, this suggests teachers not only need the opportunity, but also the right structures and organisation to support children’s autonomy, which often was not the case. This research therefore suggests that restricted environments (a single adult in one room, a lack of resources, small rooms and cramped conditions) might promote didactic methods and tight teacher control, whilst giving children greater autonomy requires adequate spaces and resources, including more adults.

5.5.6 Responding to the Tension

I have argued above that when teachers attempted to provide greater spaces for children’s autonomy within the ecosystem of teacher control, they often experienced pressures and tensions. These tensions sometimes caused teachers to struggle with the programme or to reject some of the practices. In response, some teachers restricted the strategies we had introduced and related practices, for example by only using them in some subjects (typically topic or science), with only some year groups, and only some of the time. Teachers also often described some strategies as not appropriate for their classrooms, though which strategies these were varied between teachers because of their contexts and personalities.

However, these tensions further disconnected practices relating to children’s autonomy from each other. Some teachers also responded by wanting to further limit choices for children, as they interpreted the difficulties they encountered as a sign that children needed less autonomy, rather than more. These teachers relied more on trivial choices, as a token of autonomy. Similarly, some teachers blamed difficulties on their teaching ability or on the children’s lack of skills, motivation or attitude, rather than having a diagnostic approach based on children’s needs, such as asking what might be needed to overcome difficulties or for children to feel empowered.

However, teachers (who sometimes were the same ones who struggled and at times responded negatively to the programme) also sought solutions to these tensions and difficulties. In particular, teachers emphasised the importance of being flexible and open-minded about children’s ideas, as well as creative. Beth spoke of weaving the curriculum into children’s interests in order to meet standards. Teachers also found compromises: for example, whilst Claire and Julie needed all children to complete a teacher-set task for every

lesson, they found that they could be more flexible with children by compromising on the timing of when these tasks were completed. Similarly, even teachers who struggled with the programme showed flexibility and understanding towards children in other ways – for example, Anne showed a high level of sensitivity and warmth towards children, in particular when they were distressed. Teachers also used strategies considered to support children’s feelings of autonomy, such as giving reasons and rationales when children’s preferences could not be followed. For example, when a child asked Danielle if he could join a table, she replied “You can when there's a space, there's no space at the moment” (mid-point lesson 1). Finally, despite these tensions, some teachers also experienced successes with the programme, which I now discuss.

5.6 Finding Commonalities in Successes

Despite the tensions described above, teachers sometimes reported being successful in increasing children’s ownership of their learning. These attempts were described in positive terms because either children had responded with enthusiasm or because teachers had been able to observe children’s empowerment directly.

5.6.1 High Motivation, High Capability

Children’s motivation was clearly present when teachers reported positive changes to their practice. For example, when speaking of children’s experiments, Julie reported that the children had taken responsibility for watering their own plants (“They’re having to look after them. I’m taking no control of it.”) and how children were enthusiastic about this (“You can hear them talking ... each day they’re coming in like (gasps) Has it grown?” CI interview). This contrasts with situations where teachers described children relying on the teacher – in this example, the teacher was able to relinquish all responsibility over the watering of plants both because she let children have the ownership and because children’s motivation made it possible for her to do so. Motivation was both a sign mentioned by teachers that the practice had been successful, but also seemed to be a necessary condition to the practice being successful – in other words, it was not only an outcome of greater autonomy, autonomy required high levels of motivation, suggesting a bidirectional relationship between autonomy and motivation.

In addition, as mentioned before, children needed to know what to do in order to deal with the increased autonomy and I noted earlier that children’s lack of independence was

often linked to their uncertainty or lack of confidence about what to do. Some teachers had also noticed this, noting that children were more able to cope with autonomy when it required skills or knowledge they were familiar with. For example, Beth had found that her children did not know the names of many plants. Having spent a lesson investigating different plants, she found it had helped their ability to explore the garden, “armed with more knowledge about what sorts of plants they were looking for” (Beth, final interview).

Similarly, in one lesson, Claire had asked her students to represent their knowledge about plants in whichever ways they wanted. Some children had decided to use blocks and Claire engaged them in conversations about it where they demonstrated their knowledge and she supported their decision-making by asking children what they were going to do and what they were going to use. Later on, when the teacher came back the children had added to the task by making a ‘forest’ with the blocks, and showing their knowledge not only of the parts of the trees, but also of different species and their characteristics:

The children have stuck labels to different blocks: oak, silver birch, apple, willow. Claire joins and is checking their understanding, e.g. “which part of the silver birch is silver?”, “How do we know that the willow tree is a willow tree?”. Children explain that they have put long leaves on the willow trees, red blobs on the apple trees and silver paper on the trunk of the birch trees. They have also organised the trees into ‘orchards’ (blocks to separate each type of tree). (Claire, mid-point lesson 1)

Children’s capacity to be creative with the task and to take it further was likely supported by children’s already considerable knowledge of plants, as the lesson had come at the end of a unit. In addition, they were using materials familiar to them, such as blocks and sugar paper, so they did not struggle with technical aspects of the task.

However, this is not to say that autonomy can only occur at the end of a unit of work. In Danielle’s carousel lesson on plants, children had a large amount of autonomy in their explorations of plants, and this lesson occurred early on in the term. However, most activities were accessible to all children and depended on skills they already had, such as drawing. This is not to say, either, that autonomy requires total independence. In this lesson, some activities required or benefitted from adult support – for example, print making was carried out with a TA’s help, and children’s observations and use of the microscope were supported by the teacher. But children’s independence in most activities meant that the two adults could target their support where they were needed, rather than being spread thinly.

5.6.2 *Space, not Vacuums: Stimulation and Support as Well*

So far, this analysis has been concerned with *spaces* for children's autonomy. I have discussed links with the concept of independence, both as an issue teachers mentioned when children lacked self-reliance, and as an affordance that freed up teachers to focus their support elsewhere. And indeed space is an important aspect of children's autonomy as it is seen here as the opposite of teacher control. However, looking at successful examples in the programme, clearly there was more at play than just space – teachers did not describe a vacuum with an absent teacher, as some have interpreted children's self-direction (e.g. Kirschner et al., 2006; Klahr & Nigam, 2004), but practices where the teacher was still a core element. In particular, as well as setting up the resources, the environment, and the organisation of the classroom, teachers provided *stimulation* and *support* which were essential to children's ability to make use of the space for autonomy they had been provided. Danielle's carousel lesson shows the balance of these components: children had freedom to choose activities based on their interest in the moment and space to explore and become curious, but the activities provoked their curiosity and enthusiasm through their variety and opportunities for their interest to be triggered (Renninger et al., 2018). As I highlighted above, Danielle and her TA facilitated children's activities by supporting children when they needed help and by making sure there were enough resources. For example, helping a child with observational drawing, the TA asked: "What can you see? Does it need more veins? Look at the leaf really closely" (Danielle, mid-point lesson 1). The teacher also helped children manage their time, e.g. saying to the class, "If there's something you haven't done and you want to do, you need to move on now" (Danielle, mid-point lesson 1).

Similarly, in the snail collage activity which Helen described, whilst she did not give the name of the piece, the Matisse artwork had prompted rich discussions and provided the basis for children's own interpretations. The teacher had provided the stimulus and facilitated children's conversations about it to enable their sense-making.

In addition, support may also encompass not only leading children to knowledge, but also explicitly giving information or teaching strategies when it is needed. For example, after the children discussed what some red dots could be on the leaves they were observing, Danielle looked it up and gave the correct information to the children, showing them different pictures. Different pedagogies differ in the emphasis they give to the amount of explicit teaching teachers undertake compared to a more facilitative role. In Reggio-inspired classrooms, teachers facilitate social constructions of meaning through discussions and

opportunities to experiment (Kim & Darling, 2009). By contrast, in Montessori, there may be more explicit teaching in how to do a particular task (Lillard, 2017). This relates to broader, thorny debates about the role of explicit instruction beyond the remit of this thesis, but the presence of explicit instruction in pedagogies where children otherwise have large amounts of autonomy suggests it is not incompatible with it. Our research suggests that, on the contrary, children's autonomy is dependent on children being empowered to pursue new knowledge as well as with the knowledge they have.

Before moving on, I note that there are some parallels between the notions of *support* and *stimulation* and our original concept of *challenge*, in the same way that *space* came to replace our initial concept of *choice*. In this sense, this shows there has been continuity as well as refinement over time of the ideas in the Stepping Stones project. This move away from the notion of *challenge* comes from some issues we encountered. First, *challenge* was usually misinterpreted by teachers who often understood it to mean a more difficult task (*a challenge*) or pushing children to do more or different work from what they had chosen (*challenging children*). All of these interpretations denote a teacher-centric agenda for learning at odds with children's autonomy and inner motivation for learning. Secondly, it is likely that as a research team we were not always clear on what challenge meant in practice, and that it was a general term to capture what was in fact very distinct mechanisms. For example, challenging children's executive functions might mean something very different (both in practice and in purpose) to addressing a child's feelings of helplessness faced with a difficult task, and different again to extending children's thinking through elaboration – all scenarios which would have previously fallen under our umbrella of *challenge*. By focusing on *support* and *stimulation* instead, we hope to make these distinctions clearer, both for ourselves and others.

5.6.3 Agency as Well as Autonomy

So far in this thesis, when speaking of the aims of the programme and teachers' attempts at changing their practice, I have mainly referred to children's autonomy. However, very soon after the programme ended, and as analyses progressed, members of the team independently realised – and came to the consensus – that autonomy in and of itself was not sufficient to capture what we were aiming for.

Autonomy refers to individuals' experience – typically in Self-Determination Theory – to their need to feel they are acting out of their own will (internal locus of causality),

volition and perceived choice. However, because of the focus of this thesis on the role of teachers, autonomy in our discussions with teachers was less about children's experience, and more about the situation itself – captured in this thesis through the idea of 'space'. In addition, it is easily conflated with independence (Rothbaum & Trommsdorf, 2007), which I have differentiated by its focus on self-reliance rather than self-determination. Both space and independence ignore the crucial role that teachers play when they are not being controlling, which I have described here as *stimulation* and *support*. Others have described this role as structure, but whilst structures and organisation can be designed in a way that support children's participation and autonomy (e.g. councils, organising resources for children to access them independently), other descriptions of structure can be controlling (e.g. having firm expectations, and giving information for how to satisfy those expectations, Hospel & Galand, 2016).

In addition, a conception of autonomy centred on children's experience has more to do with whether children feel they are acting out of their own volition and internal locus of causality (as opposed to because others want them to) than on the external situation. This means focusing on the extent to which children experience control and choice over classroom activities, the extent to which children endorse these and engage willingly with them, and whether they have the freedom not to engage (i.e. consent) or to have respite. Nonetheless, the focus on autonomy ignores whether this results in any increased participation, whether in action (which can be observed) or as feelings of being *part of* rather than *done to*. Therefore, long before these analyses were completed, our team had moved away from autonomy and become instead interested in the concept of *agency*.

Whilst the term has different meanings in different fields, such as sociology (Archer, 1995; Emirbayer & Mishe, 1998), educational theory (e.g. Biesta & Tedder, 2007; Manyukhina & Wyse, 2019), pedagogy (Barton & Tan, 2010; Rainio & Hilppö, 2017) and psychology (Reeve & Tseng, 2011), we believe it has more potential to fully capture the idea that children must have opportunities, motivation and capacity to act (or not to act). Central to our understanding of agency is its definition as an internal experience under the influence of the context and situation, rather than simply the affordances of the situations themselves – in other words, an individual is only agentic if they *feel* agentic. This understanding of agency is the product of a team effort to clarify the concept of agency in the context of early education and its relationship to playful learning, motivation and self-regulation (Baker et al., 2021).

There is nonetheless some overlap with the concept of *feelings of autonomy* in terms of volition and locus of causality, and indeed the term agency is sometimes used in conjunction with autonomy (Cremin & Chappell, 2019; Ryan, 2017). However, in our view research on autonomy support does not fully acknowledge situations where students might experience autonomy but not agency. For example, a few studies have found that gender stereotypes can constrain children's participation in specific play events, despite an obvious desire on the child's part to be included (Rainio & Hilppo, 2017; Wood, 2014). Another study found that even in democratic settings aiming to foster students' participation, some voices may become dominant over others (Wilson, 2015). Focusing on agency may help teachers, for example, attend to power dynamics and to barriers to children's participation and empowerment, in ways that framing practice around autonomy may not. It may also help move the focus away from independence and make it easier to see how certain situations might make it difficult for children to feel empowered to take charge of their learning – for example, as we sometimes found in the programme, where children are given space, but not support or stimulation, or they don't have the right skills and knowledge, or where learning activities are not meaningful to them. These issues relate to our focus on supporting the need for autonomy, but not specifically addressing the needs for competence and relatedness described in Self-Determination Theory. *Support* might allow children to feel agentic not only because it provides them with skills and knowledge, but because by feeling supported children also feel a sense of competence in their work as well as safety in trying out unfamiliar tasks, which would itself be under the influence of feelings of belonging and relatedness. This highlights the inter-relatedness of these different needs and the importance, in the future, of considering all three needs described by Self-Determination Theory together.

5.6.4 Power and Control

Moving away from autonomy and towards agency requires paying greater to power and control. However, the notions of power and control are themselves fraught with difficulties. By encouraging teachers to share power and control in the classroom, this implies that the teacher relinquishes some of it. However, it is sometimes assumed that giving individuals power and control means *complete* power and control. It is also sometimes extended to mean power over others – for example, when giving children more power, concerns are raised that they will use that power over other children. In addition, there is also a concern that giving up power and control necessarily results in becoming powerless or out-

of-control – and therefore that by asking teachers to increase children’s autonomy, agency and power, we are asking them to be powerless, or at the mercy of children’s will. These issues stem out of an absolutist vision of power and control – that power is diminished when it is shared, as though there is only so much to go around. To illustrate how sharing power and control is more nuanced than this, I draw on two examples from the Stepping Stones programme.

In Anne’s lesson where she felt ‘out of control’, her relinquishing of control did not necessarily increase children’s control or power. At best, the children likely had fun, but were not necessarily empowered to take their learning further: they still acted mostly passively, attempting to follow the fair test procedure the teacher had introduced but with little success; some children used their freedom to procrastinate. So although Anne had felt she had completely lost control of the lesson, it seems unlikely from the children’s response that they experienced a large sense of empowerment, let alone complete power. There was not a direct transfer of power from the teacher to the class.

By contrast, in Beth’s series of lessons on plants, Beth still very much remained in control. She described the process of letting go of power as scary, as she did not know where her lessons were going next. But she did not describe this experience in terms of powerlessness or feeling out-of-control in the way Anne had. In fact, Beth was excited about it. However, she described children’s enthusiasm in ways that suggest the children had indeed felt greatly empowered, because the ideas for what they were investigating had come from them. In other words, Beth had likely a much larger impact on children’s feeling of empowerment compared to the ‘amount’ of control she had lost.

What these examples suggest is that, whilst it is true that sharing power and control entails that the dominant party give up some of their power and control, it does not necessarily follow that the power is directly transferred to others, as though power and control were a currency that could be exchanged between bank accounts.

5.7 Alternative Explanations: The Paradigm Shift and the Balance of Methods

This chapter has put forward a model to explain teachers’ attempts to increase opportunities for children’s autonomy in the early Primary classroom in England, drawing on the experiences and observations of teachers taking part in the Stepping Stones programme. I have argued that classrooms operate as an ecosystem of teacher control, i.e. a complex web of inter-dependent practices where the teacher is the centre of power and knowledge. Yet *at*

the same time, teachers sometimes provided spaces for children's autonomy in various ways and amounts, but through discrete rather than interconnected practices. I have provided evidence for how ubiquitous and inter-related the practices of teacher control were, and how they were driven by high-stakes accountability, a performative mindset and a focus on specific learning objectives. I have also shown that teachers often found it difficult to increase how much space they gave to children's autonomy because of the pressures they faced and because the co-occurrence of these ways of working created dissonances and tensions, which sometimes resulted in issues for the teachers and for the children. However, I also showed that teachers were sometimes able to negotiate these tensions and could be successful in increasing children's ownership of their learning, even within the ecosystem of teacher control.

In this section, I wish to further evaluate this explanatory model by assessing other ways that researchers have conceptualised the tension between teacher control and children's autonomy in the classroom – in other words, to argue that this model explains the evidence better than existing models (Figure 8). In particular, the literature tends to coalesce around two main models. The first is the idea of a paradigm shift: because practices are interdependent, and because teaching for children's autonomy is dependent on beliefs and classroom cultures that are congruent with it, teachers cannot simply tinker with their existing practice but need to operate in a completely different teaching paradigm. This is for example the position taken by Lillard (2019), who argues that Montessori practice is incommensurable with the state education system. The present research provides some evidence in favour of this theory. Teachers experienced tensions between different ways of working, and sometimes giving children greater autonomy failed because the broader, surrounding ecosystem of teacher control as well as teachers' conflicting goals and beliefs. The interconnectedness of practices and norms in the classroom therefore made it difficult to make successful small changes. However, it was not impossible, as evidenced by the successful examples we have described. This suggests that children's autonomy is in tension but not incommensurable with the rest of those teachers' practice. It may therefore be that the more teachers' practice is coherent towards greater autonomy for children, the fewer difficulties they will face, but this research also suggests that teachers do not need to operate in a different paradigm in order to begin making space for children's autonomy and agency. Even teachers in the most-controlled classrooms were able to make small changes to their practice.

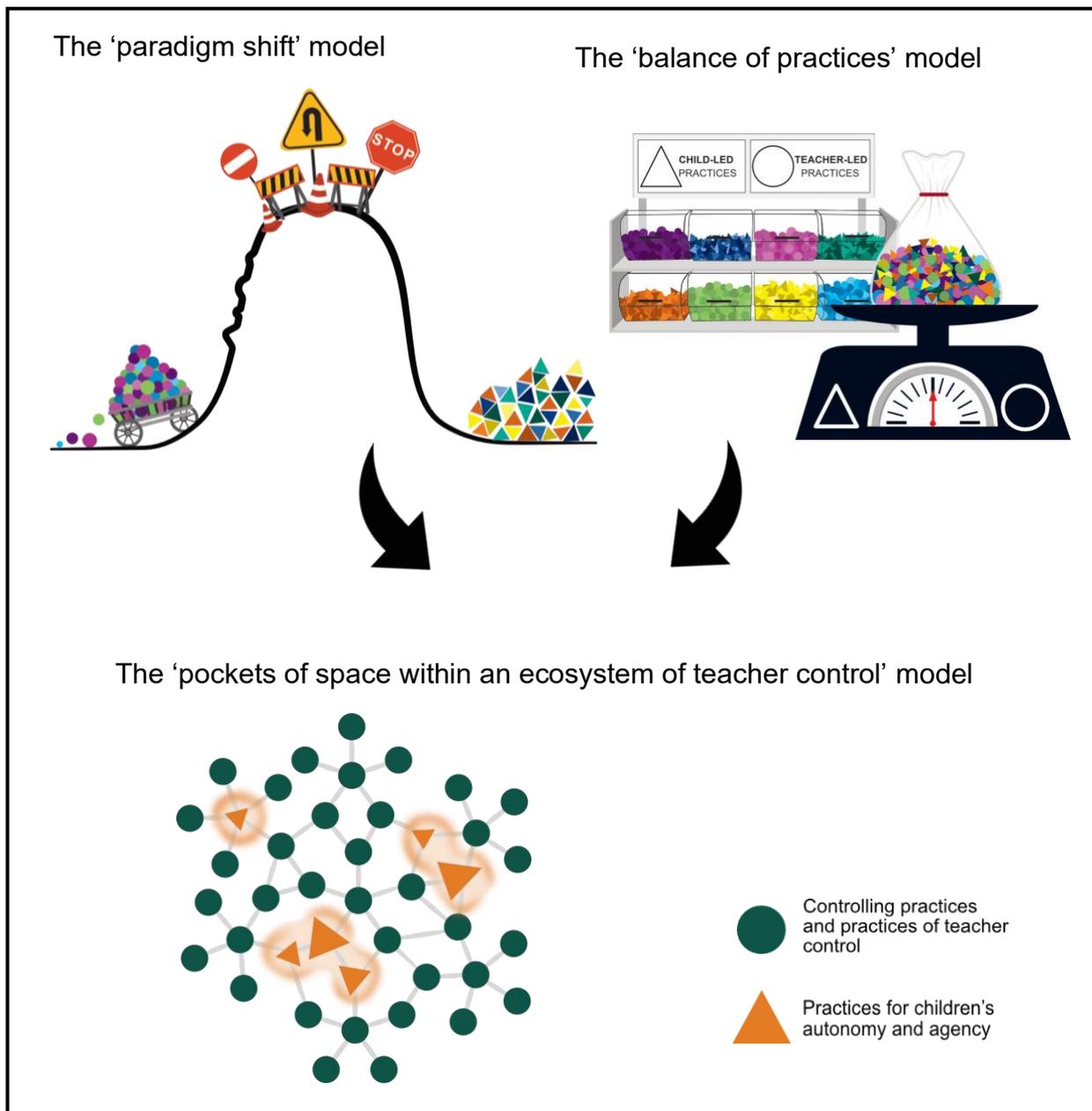
The second way in which the opposition between practices of teacher control and for children's autonomy has been described is in terms of a 'balance' of practices or as 'hybrid'

teaching. This may be a balance between student-centred or teacher-centred methods, or a balance between child-led free play and teacher-led instruction. This conception is problematic in a number of ways. First, it does not acknowledge the inter-connectedness of practices in the classroom. This means that it cannot deal with the ways in which different practices reinforce each other, or how giving children more autonomy might conflict with the dominant way of working in the classroom. Therefore by putting these different practices at different ends of a pole or along a spectrum, it does not allow enough room to understand the ways in which these practices are co-occurring and conflicting. Other research has also shown that structuring practices and autonomy-supportive practices are not mutually exclusive (Jang et al., 2010; Tzuo, 2007), and the present research shows that these interactions are complex, with teachers navigating a number of tensions.

Secondly, the ‘balance’ model does not make room for questions around empowerment and agency, and whether students endorse activities or the role they have been attributed – in other words, it focuses on whether activities are child-directed rather than child-centred (Pyle & Daniels, 2017). For example, Zosh et al. (2018) describe a spectrum of play and learning (2018) depending on whether the child or the adult initiates or directs the task; the linearity of the model implies a conception of control where it is either the teacher’s or the child’s. Often, these practices are in fact better described as methods or strategies, with researchers arguing for a ‘pick’n’mix’ approach whereby teachers pick whichever method (whether child-centred or teacher-centred, child-led or teacher-led) that best fits their aim for a given lesson (Bremner, 2019; Jensen et al., 2019) – a position which has sometimes been described as ‘learning-centred’ (Bremner, 2019) to mark its distance from the ‘false dichotomy’ of child-centred versus teacher-centred teaching. In other cases, a hybrid approach is recommended in order to meet diverse educational goals (Pascal et al., 2019; Siraj-Blatchford et al, 2008; Zosh et al., 2018). However, this overemphasises adult-imposed learning objectives, which as we have seen are a main driver for restricting children’s autonomy. In this sense, there is no consideration of what the child’s agenda might be, but only whether the methods fit the purpose the adult has decided. In addition, this conceptualisation does not make space for how control and autonomy are experienced – that is to say it places the emphasis on outward control rather than on internal locus of causality. However, the same action may be interpreted differently by different children, which may be different depending on the cultural context (Lee et al., 2018) as well as between individual children and situations (Baker et al., 2021). It also does not account for situations where

children may feel empowered even when teachers provide a large amount of structure and guidance (Tzuo, 2007).

Figure 8 Visualisations of Contrasting Models of Teacher Control and Children’s Autonomy in the Classroom



Note. Graphics by Allison Haack based on the ideas presented in this thesis and preliminary visuals provided by the author. Used with permission. The ‘paradigm shift’ model represents the idea that it is not possible to move from one way of teaching to the other incrementally. The ‘balance of practices’ model represents a ‘pick ‘n’ mix’ conceptualisation of practices based on adult goals.

Whilst in some cases the call for a balance is in response to a shift towards more teacher-directed time (e.g. Pascal et al., 2019), there is a risk that the language of ‘balance’ or ‘hybrid’ practices encourages ‘middle-ground fallacies’, which would suggest somewhere in between those extremes in the correct place to anchor practice (e.g. Skene et al., 2022). However, what the ‘right’ balance should be is unclear, as it depends on many factors, not least the desires of the children involved in individual situations and the values and beliefs about education that adults bring.

By contrast, in this chapter I have proposed considering teachers’ practice as an ecosystem of teacher control but in which teachers can make space for children’s autonomy and agency. This view of the tension between teacher control and children autonomy is therefore in line with ‘transactional’ models of learning, where the respective roles of students and teachers are under constant negotiation, rather than opposite poles (Elen et al., 2007), as well as with dialectical conceptions of this tension which refuse simple binaries (Rainio & Hilppö, 2017).

This allows us to not only portray the complex dynamics of classrooms with more nuance, but also appreciate the tensions teachers experience through the co-existence of conflicting ways of working as well as the ways in which they negotiate these tensions. It can also allow us to make a case for children’s autonomy which is not dependent on prescribing specific teaching methods, but instead asks us to engage with what practices mean for children’s empowerment and agency.

5.8 Conclusion to Chapter 5

In this chapter, I have argued that teachers function in an ecosystem of teacher control, underpinned by institutional structures and pressures. At the same time, teachers provide pockets of space for children’s ownership of learning and their participation, but these are small and disconnected. As a result, teachers experienced many tensions and difficulties in trying to provide children with greater autonomy, and the changes they made to their practice were small, despite many teachers’ enthusiasm for the programme and its goals.

Because of the focus of autonomy on self-direction and even sometimes its conflation with independence, our team now focuses on the idea of agency. We feel agency more fully captures the emergent experiential nature of children having control, participating in and willingly endorsing learning activities. It also better highlights the role of the teacher, for example in providing stimulation and support as well as space, and therefore in their need to

be responsive to what is happening in the moment. We have also moved away from the concepts of choice and challenge, which were often interpreted by teachers as meaning, respectively, limited choices between options and imposed difficulty, neither of which are likely to lead to enhanced experiences of autonomy or agency.

The ideas that children need not only space, but also stimulation and support, and that teachers need to be responsive to the individual children in their classrooms are not new. They have not been ‘discovered’ through this work, and they can be found in many existing pedagogies and education research – indeed some teachers already used some of these practices before they took part in the Stepping Stones programme. However, by highlighting these aspects of practice as well as focusing on the notion of agency, our hope is that it will help teachers view their practice through a different lens than that provided by the teacher control ecosystem.

Chapter 6 - Part I Discussion and Conclusion - Supporting Teachers' Provision of Autonomy and Agency in the Classroom

In this chapter, I conclude Part I by considering how we might be able to better support teachers' provision of autonomy and agency in the classroom, given the difficulties highlighted above. This requires considering the ways in which the Stepping Stones programme itself supported teachers' change in practice, as well as what may have been missing. Whilst investigating teachers' learning processes in general was not an explicit aim of this research, these are relevant to understanding the limitations of the present research, and therefore what the programme might need to do better in the future. Specifically, I focus on two areas where teachers might need to be better supported: in their capacity to take an inquiring stance towards their practice, and in the skills they need to develop to teach in a way that supports children's feelings of autonomy and agency. Finally, I also discuss some further limitations of the research described in Part I.

6.1 Developing an Inquiring Stance Towards Changes in Practice

6.1.2 Differences in How Teachers Dealt with Change and Difficulties

All teachers reported enjoying their participation in the Stepping Stones programme, in particular the chance to come together to share practice with other teachers, but their enthusiasm for the idea of children's choice and autonomy varied. Whilst some teachers reported successes, other struggled more with the tensions they experienced. Whilst many of the tensions were caused by the practices themselves, there were suggestions that teachers' disposition towards changing their practice also caused issues.

For example, whilst some of the teachers saw the programme as an opportunity to try something new and "think outside the box" (Lisa, final interview), a few of the teachers struggled much more with the programme because the practices deviated from what they were comfortable with. Both Anne and Helen described the practices we had introduced as going against practice which they knew "worked". They therefore questioned the value of the programme ("I know it works. So why change it if it already works" Helen, final interview; "I know the kind of things really, in my heart of hearts, that would work", Anne, final interview.)

However, this way of thinking about their practice, as something fixed that could not be improved, could make it difficult for teachers to notice what could be done differently. Whilst it may have been partly a reaction to the practices we were encouraging specifically, the way in which these teachers talked of ‘what worked’ suggested it was more general reluctance to change practice. Anne made the point that she was open to change, but immediately countered this by adding that because “children don't change” there were set practices that always worked and others that did not. She also cited “teachers’ experience” as one of the barriers to the programme, adding that new teachers may be more open to new practices. In parallel, Anne was particularly critical of herself and expressed a lack of confidence in her ability to try the new practices (“It didn’t work well with them. But that may have been me” CI interview; “I know I didn't achieve what good examples of flexible planning are” final interview). This likely contributed to her returning to her previous practice.

By contrast, Helen grappled with the programme and found it difficult, but at other times described this process in positive terms (“I did really enjoy it. I really found it a challenge ... a positive challenge”, Helen, final interview). In particular, she mentioned “loving” being able to try something new and reflect on what they would do next. Nonetheless, she also described experiencing a large amount of “friction” (Bakkenes et al., 2010, p. 539) between the what she wanted to achieve and what happened in practice. Helen and Anne could therefore be described as having problematic learning experiences (Vermunt et al., 2019) because of the intensity of the difficulties they encountered and the grief it caused them.

By contrast, whilst Lisa reported that when things went “wrong” when trying a new practice it made her not want to try it again, she also felt that it was a learning opportunity (“how could I make it right next time”, Lisa, final interview). Therefore the very process of the programme – centred around teachers’ own inquiries in their classrooms – seemed to generate difficult feelings for some teachers, whilst others embraced it. Other research has shown that encouraging similar shifts in practice (e.g. towards enquiry-based learning) challenges teachers’ existing routines and their professional narratives regarding why they teach (Williamson & Morgan, 2009).

It may be that in the future we need to pay greater attention to how we nurture and support a risk-taking, inquisitive stance to practice rather than assume that the opportunity to do so will be embraced by all teachers. It seems to be this inquiring stance that allowed some teachers to approach difficulties and tensions without judgement and to work through lessons

that did not turn out as they had expected. Indeed, Opfer and Pedder (2011) argue that dissonance (or “cognitive conflict”) can in fact be a force for change, to allow teachers to ‘unlearn’ what they believe and know, in order to learn new practices. Teaching for autonomy necessarily involves dissonance, because it is so at odds with the ecosystem of teacher control. Indeed, we could wonder whether Julie and Claire experienced sufficient dissonance – their change in practice had been imposed by their head teacher, and therefore it had not been accompanied by a questioning of power dynamics in the classroom. It had practical implications, but not ideological ones. However, as seen in Anne’s example, too great a dissonance led teachers to return to practices that felt comfortable (Coburn, 2001). In the next sections I consider how this inquiring stance might be better supported in the Stepping Stones programme by drawing on the wider literature on teacher learning.

6.1.2 Supporting Teachers’ Meaning Making

An inquiring stance supposes that teachers encounter difficulties (such as dissonance or problematic learning experiences) as opportunities to learn and wonder about. The co-research process described in Chapter 3, with teachers tinkering with practice rather than conducting their own research, may not have been successful in fostering this kind of mindset. If we understand inquiry as solving a problem or deepening understanding (Cassidy et al., 2008), then teachers’ inquiries should be centred around concerns or questions teachers want answers to (Korthagen, 2017). In this sense, teachers are not simply experimenting or testing practices, but posing their own questions. For example, the challenges teachers encountered could have been starting points for such enquiries, rather than simply obstacles. Supporting teachers’ ownership of inquiries may also require us to involve teachers in assessing the impact of their practice, for example through documentation of children’s learning (Reggio Children & Harvard Project Zero, 2001) or the development of assessment tools (Perry et al., 1999).

In addition, we could question the extent to which teachers engaged in analyses of their practice. In this thesis I have not formally analysed teachers’ conversations for learning processes (e.g. as in Vrikki et al., 2017), but it seems they were centred on agreement and descriptions rather than productively analysing why things had unfolded as they had and what might have been going on with children. For example, teachers were often decisive in their judgements of what children could do or not do, and could or could not cope with. This learning pattern has been called application-oriented as it is focused on applying what has

been learned through experimenting with practice but staying with the boundaries of their existing theory of practice (Vermunt et al., 2019). By contrast, meaning-directed learning is focused on trying to understand why certain methods work and why students respond in certain ways and involves co-construction of knowledge, which is characterised by joint problem solving and short turns of talk where each individual develops an idea by building on the contributions of others (Vermunt et al., 2019). Encouraging meaning-directed learning and more collaboration between teachers might help them engage more deeply with the difficulties of providing children with greater autonomy.

It may also be that we need to encourage greater reflection after practice, which has been shown to help teachers change their beliefs (Tillema, 2000). Although we held debriefs after lessons, the sessions were brief and not aimed at reflection; we did not hold a mirror to teachers' practice or ask questions that would have led teachers to challenge their interpretations of events. It may therefore have been that teachers made quick evaluative judgements based on their prior experiences rather than systematically scrutinising information (Gregoire, 2003), as might be the case by the distance and reflection afforded through documentation.

Helping teachers be more meaning-directed and ask questions of their practice rather than simply trying to 'fix' problems may be addressed through changing aspects of the workshops; for example, better facilitation may help to make teachers' assumptions or beliefs about teaching more explicit or to encourage dialogue. Sharing practice through stories and materials they are already working with such as lesson plans (Oliver et al., 2018) could provide opportunities for teachers to share what they have done and reflect with others, perhaps in smaller groups without the 'gaze' of the research team. Although teachers had begun to do this in the workshops, this could be more explicitly part of the programme.

It may also be that additional methods would be helpful in framing teachers' inquiries. Teachers' experimentations and reflections were unguided outside of the facilitated workshops, and teachers may have benefitted from more explicit frameworks so that they could conduct inquiries of their own. For example, models of inquiry such as guided reflections (Korthagen, 2017), action research cycles (Baumfield, 2008), spirals of inquiry (Timperley et al., 2014), lesson study (Lewis et al., 2006) or coaching centred on dialogue around authentic challenges and dilemmas (Lofthouse, 2019) may help teachers approach specific problems with an inquiring frame of mind. Korthagen's model of self-guided reflection (Korthagen, 2017) provides concrete support for teachers to reflect on their own behaviours and their sources, and then devise alternative methods in answer to these

reflections and trial them. Specifically, reflections focus not only on rational analysis of practice but also on the emotional and motivational aspects of lessons, which may allow teachers to link their actions to underlying beliefs. Similarly, Baumfield (2008) describes integrating the ‘plan, do, review’ model with cycles of inquiry which emphasises the posing of problems and hypotheses, which are then evaluated in the review phase. This would not only provide teachers with greater ownership over what they tried in their classroom, but would also help make more explicit teachers’ own theories about what they were attempting by trying different strategies. The ‘Spirals of Inquiry’ framework proposed by Timperley and colleagues (2014) also supports teachers in identifying a need in their school and take action to address it by providing concrete tools and steps. For example, it asks teachers to reflect and develop ‘hunches’ about what is going on for their learners, including by involving children in discussions about their learning and their experiences, which would make this framework particularly well suited to our programme aims. Teachers are then encouraged to seek out learning opportunities to allow them to take action. Lesson study might help teachers feel that change is possible and the collaborative approach to planning and problem solving may help reduce problematic learning (Vermunt et al., 2019). In addition, both lesson studies and peer coaching can make use of videos (e.g. Vermunt et al., 2019; Lofthouse & Thomas, 2017), which might make it easier for teachers to reflect on specific events.

6.1.3 Supporting Teachers’ Self-Efficacy

It also seemed that teachers’ own feelings of safety and confidence were important to help them embrace a tinkering approach to making changes to their practice. Supporting teachers’ self-efficacy has indeed been recognised as an important aspect of professional development (Lofthouse, 2019; Reeve & Cheon, 2016). In particular, Gregoire (2003) theorises that self-efficacy is essential to whether teachers perceive change as a challenge or a threat. This suggests that we may need to pay greater attention to teachers’ self-efficacy in the future, in particular because giving children more autonomy can erode teachers’ confidence and sense of control over lessons, as was the case with Anne.

Self-efficacy can also be supported by positive relationships (Gregoire, 2003). Both coaching and communities of practice require relationships of trust and safety, characterised by dialogue and reciprocity (Lofthouse, 2019; Lofthouse & Thomas, 2017). However, building trust and relationships requires time, and therefore it may be that the programme’s short length hindered this aspect of teachers’ learning (Darling-Hammond & Richardson,

2009). Coaching and collaboration between teachers may be particularly helpful to foster self-efficacy confirmation; anchored in ethics of care, it involves identifying a better self and encouraging its development (Flint et al., 2011). Playful and joyful approaches to teacher inquiry (e.g. Baker & Ryan, 2021) may also help foster teachers' curiosity and feeling of safety, as a more playful mindset can help teachers take risks and experiment with practice. For example, some researchers have used playful provocations and playful ways of sharing research (e.g. creating books, games, and interactive demonstrations) to help teachers experience wonder and delight in their practice.

Common across these approaches are the need for a non-judgemental space, trust, curiosity on the part of the teacher, and time. It may have been that although we aimed to create a space for teachers to experiment with practices – and teachers' reports suggest we did – the top down nature of the community of practice and its short timeframe may have made it difficult for teachers to develop the relationships necessary for this kind of approach to fully realise its potential. Oliver et al. (2018) argue that “top-down” learning communities may result in no actual community in the sense that participants go their own way once the programme has ended, and this may leave no time for the community to mature. Other than for teachers who taught in the same school, this was certainly the case here. We could question the extent to which teachers worked collaboratively – that is to say together rather than simply with – through the community of practice. However, top-down approaches can be helpful in some ways, such as providing facilitation and funding as well as professional support (Tarnanen et al., 2021; see also e.g. Hauge, 2019). In addition, ‘bottom up’ communities can also face problems. For example, different teachers within the same community may hold contradictory mental models, e.g. regarding school rules, curriculum or the role of students (e.g. Tarnanen et al., 2021), which means communities may not be cohesive. Collective participation can also inhibit innovation when the majority in the group resist change (Sherman & Teemant, 2021). Nonetheless, with respect to our programme, this suggests we need to give greater consideration to the quality of the community and teachers' feelings about it.

6.1.4 Understanding Teachers' Beliefs

Many frameworks of teacher learning state the importance of teacher beliefs for the success of professional development (e.g. Desimone, 2009; Opfer & Pedder, 2011). Indeed, some argue that changing teachers' beliefs is an important aspect of these programmes.

Beliefs are defined as conceptions, personal ideologies, world views and values that shape practice and orient knowledge (Aguirre & Speer, 1999). In the present research, relevant teacher beliefs included those about the nature of learning, the role of the teacher in children's learning, and the place of teacher authority and power in the classroom as well as the desirability and nature of children's autonomy.

Research suggests that beliefs and knowledge are more readily modified than practice (Bakkenes et al., 2010), or theorises that the former need to be modified before the latter (Desimone, 2009). However, others argue that there is a reciprocal relationship between beliefs and practice, with "changes in one being contingent on changes in the other" (Opfer & Pedder, 2011, p. 395). In our programme teachers implemented changes in their practice without signs that their underlying beliefs had been changed – though in some cases, they had been reactivated after years of not being able to act on them ("knowing that it's the right thing, and it's always been the right thing", Helen, final interview). In fact, in our programme we found it difficult to change beliefs and knowledge, which meant that changes in practice were not *deep* – they rarely resulted in questioning deeper questions around power and control in the classroom.

In addition, as we have noted before, teachers' beliefs about children's autonomy were sometimes contradictory. Other research has shown that teachers' practice and beliefs towards child-centred practices are not always consistent (e.g. Sak et al., 2016) and beliefs play an important role in teachers' selection and prioritisation of the multiple goals they may hold for children's education (Aguirre & Speer, 1999). For example, Helen's struggle with the programme could be partly attributable to the different, contradicting beliefs she held about the value of child-centred practices versus what she had been told to do "for years" (final interview) and had therefore likely internalised to some extent. Similarly, Anne's struggle with the programme was likely partly due to the fact that she found the idea of children having more control uncomfortable ("it also makes you feel like the children are dictating ... where they're going to end up ... I just feel a bit uncomfortable with it" Anne, final interview). Teachers' beliefs can be expressed as 'obligations' to teach in particular ways. For example, Webel and Platt (2015) describe a mathematics teacher who had a goal of increasing students' agency, but this was constrained by obligations she felt around mathematics teaching, such as the need for students to develop efficient methods of problem solving such as a subject-specific norms.

In future research of this kind, focusing on understanding teachers' beliefs may be particularly important as they were often implicit, and it may require a lot of work to bring

them to light. However, it is questionable whether it should be the aim of a programme to change teachers' beliefs unless teachers hold beliefs that are demonstrably false or damaging. In the case of our programme, where increasing children's autonomy is not only underpinned by evidence but also by a set of non-neutral values, it would seem more appropriate to increase teachers' knowledge of theories and to help them understand contradictions in their beliefs. The practice of inquiry and reflection we have described above may therefore support teachers in uncovering underlying tensions and contradictions in their own beliefs and practice. Nonetheless, it may be that the reluctance of some teachers to engage in the programme – like Anne – is better understood in terms of a different set of beliefs and values. Therefore, it may be that the programme is simply not appropriate for teachers who do not value children's autonomy, both as an educational outcome and as an educational experience. As Opfer and Pedder (2011) argue, teachers' beliefs influence not only the decisions they make in the classroom, but also what they are willing to learn.

6.2 A New Set of Skills

Anne mentioned that she did not have the right skills for the sorts of practice we were encouraging. Though in her comments she suggested that this was fixed ("I'm not very good at it," Anne, final interview), it opens up an interesting proposition we had not explicitly considered: that teaching for autonomy requires not only support for teachers (in the form of the Community of Practice and ideas to try), but also specific skills to enable these practices to be implemented successfully.

6.2.1 Responsiveness and Flexibility: A Dance

Many of the strategies we encouraged rely on teachers being able to respond to situations as they arise, rather than scripted or pre-planned procedures. Because agency is an emergent internal experience, it is entirely contextual. There cannot be a recipe for teachers to follow, or a set of rules they can apply, because the same teacher actions may or may not lead to agency depending on how they are received in any given situation. As Loris Malaguzzi (the founder of the Reggio Emilia approach) notes:

There's a difference between the environment that you are able to build based on a preconceived image of the child and the environment that you can build that is based on the child you see in front of you – the relationship you build with the child, the games you play. (Malaguzzi, 1994, p. 1)

The teacher must therefore constantly adapt to what is happening right then in their classrooms. Claire indeed spoke of the need to be flexible and creative, not only to follow children's interests, but to strike this balance ("It's just getting it right ... we have to be ready and prepared for the changes of days and lessons," final interview).

In this perspective, the strategies we provided were only starting points that needed to be constantly tailored, and their meaning in any given situation must be constantly questioned. For example, when being flexible in her planning, Beth had to make a number of decisions as to how she would trigger children's questions, which questions she would use and which she would not be able to fit in, how she would weave in the curriculum, what support children needed in setting up experiments, what discrete teaching she might need to include, how she would support discussions, etc. – none of which could be stipulated in advance, as they needed to meet the needs and responses of the real (rather than theorised) children in Beth's classroom.

In this perspective, there is also therefore no definite, predetermined amount of support or space that must be given in order for children to feel agentic – what *space*, *support* or *stimulation* might look like in practice therefore never stands still. These moves between stepping in and letting go have sometimes informally been qualified as a dance, as adult and child are on an equal footing, with the adult and child constantly responding to each other's moves. However, engaging in this 'dance' is skilled work, precisely because there cannot be a set of rules to follow.

6.2.2 Early Years Teaching Skills

Perhaps because many of the practices we encouraged came from Early Years pedagogies and research, it is possible we had not sufficiently considered how teachers in older classrooms might not have acquired these skills through their teacher training. Other research also supports the idea that different ways of teaching may require of teachers a "fundamental change in their pedagogical role" (Bakkenes et al., 2010, p. 534). For example, Bakkenes et al. (2010) argue that teachers aiming to foster self-regulated student learning need to adopt different roles (e.g. "diagnostician" and "challenger"), reflect not only on students' learning but also on their learning processes, and adopt a large number of new teaching skills, such as teaching metacognitive strategies and coaching students in their use. This is potentially only a small part of these teachers' struggle with the programme – Anne taught Reception as well as Year 1, and therefore presumably had skills necessary for

teaching Early Years – but could be a fruitful avenue of investigation for our programme. For example, following children’s lead might require teachers to ‘plan in the moment’ (Ephgrave, 2018), as well as knowing how to interact with children without interfering (Fisher, 2016). It would also depend on teachers’ ability to attune and listen to children (Rinaldi, 2001; McNally & Slutsky, 2017). Engaging children in sense-making would require an understanding of sustained shared thinking as well as how to facilitate dialogue (Gripton & Knight, 2020; Purdon, 2013; Siraj & Asani, 2014). Creativity and knowledge of how to create and sustain interest might also be needed (Renninger et al., 2018). Creating the sorts of classroom cultures that support autonomy and agency might also require teachers to develop greater sensitivity (Rimm-Kauffman et al., 2002), as well as the ability to support children’s socio-emotional development (Pakkarinen et al., 2020).

These are teaching skills that may not have been needed to the same extent in classrooms focused on tightly-sequenced learning, rote practice and compliance. As Helen pointed out, “actually knowing how to do it is difficult” (Helen, final interview), and teachers may need help in learning these new skills.

6.2.3 Helping Other Members of Staff to Develop New Skills

Teachers also noted their TAs sometimes lacked those skills. Julie and Claire, in particular, had struggled to help their TAs move away from their traditional role towards a facilitating one (“they want to do everything for [the children]”, Julie, Workshop 3; “They like it to be structured, the TAs” Claire, Workshop 3). Other teachers also noted this (“it’s training those staff to have the skills to sit back a bit” Danielle, Workshop3; “They expect there to be an outcome ... they want to know what it should look like at the end, as opposed to the process of getting there” Stephanie, final interview).

In order to support their TAs, Claire and Julie had modified some of our resources to provide focus points for their TAs during lessons. For instance, they had given a TA the strategy of ‘observe, listen and notice children’s engagement’, and provided a structured sheet with a focus activity (e.g. halving pizzas), assessment questions (e.g. ‘Can they make equal shapes?’) and guidance (e.g. ‘Allow children to explore and question each other’), as well as a space for TAs to note down their observations.

This suggests a need to consider not only teachers but all adults in a classroom, as promoting children’s agency appears to disrupt not only the traditional role of teachers, but of all members of the classroom.

6.3 Further Limitations (Part I)

6.3.1 Limitations of the Research Design

The limited time of the programme (10 weeks) not only likely had an impact on teachers' ability to adopt new practices, but it also had an impact on data collection. The short timescale was compounded by the relatively large number of teachers (data collection was carried out in 11 classrooms for initial and mid-point visits and interviews, and nine classrooms thereafter), which meant we spent little time in each classroom (6 lessons per teacher). Whilst this was beneficial for investigating differences and similarities across classrooms and teachers, it limited the amount of time we could spend in each classroom. As a result, many practices between observations were reported by teachers rather than directly observed, which could have affected "descriptive validity" (Maxwell, 2012a).

6.3.2 Interviews Limitations

There were a number of issues with the interviews which may have affected the quality of the data. First, the interviews varied in length from 15 minutes to 1 hour and may have benefited from being longer. However, the reality of working with teachers is that they are short of time and so interviews often took place at lunch breaks or after school under time constraints. This meant it was not always possible to follow up on all threads of the discussions. In addition, there were unavoidable differences between the two researchers who collected the data. For example, my running records tended to be longer than those of the other researcher. There may also have been differences in our interviewing styles, which may have affected teachers' responses. It is also likely that our different understandings of the Theory of Change altered what we noticed. For example, it is possible that my own interest in the use of rewards and punishments, and the use of controlling or autonomy supportive language, meant that I was more susceptible to noticing and recording this type of interaction than the other researcher.

6.3.3 Children's Voice

Although the role and voice of the teacher was given much consideration in the programme design, the role of children's voice was relatively absent. At this stage of development of the programme, we were not looking at the impact on children per se (though it did emerge through the teachers' own reflections), and therefore the children were not

consulted in the research. In addition, involving children's voice would have required more than simply treating them as an additional source of information (see Chapter 7 for further discussion). Instead, children's voice can be seen as synergistic with questions of choice, empowerment and agency. Therefore, we had hoped that children's voice would be embodied wherever teachers tried to consult children, changed the power structure of their classroom, and allowed children the chance to participate in decisions about their learning. The intention was that this would result in more genuine opportunities for children to have a say over what happened in their classroom than 'children consultations' by the research team could hope to produce.

However, in reality deep questions about children's rights to voice and participation (rather than simply choice and independence) were not fully explored in the workshops. As we have seen, the changes teachers made to their practice were small, and the children may not have indeed been aware that their teachers were attempting anything different from their usual practice. In particular given the importance of children's *experience* of autonomy and agency, in the future we should aim to give a more prominent place in the research design to children's voice and participation in the research.

6.3.4 Uncertainty of the Impact on Children.

At this stage of the research, we were concerned with teachers' practice rather than with the impact of the programme. However, the impact the practices had on children was intertwined with teachers' attempts, as teachers' engagement with the programme was linked to children's responses to the new practices they were trying. Therefore, children's responses became essential to our understanding of what teachers felt they could do; for example, teachers discussed practices in terms of whether children had been engaged or not, whether children had become disruptive or had found activities difficult, as well as on the quality of their learning. However, children's responses were often reported by the teachers rather than observed, and therefore in the future it will be important to investigate directly the impact that different practices had on children. This entails not only understanding impact in terms of general outcomes, but in terms of children's *responses* to the practices in the moment, which suggests a more contextual understanding of impact. Our investigations of a measure of children's depth of engagement in Part II may help with this.

6.4 Conclusion to Part I

Teacher learning takes place in the connection between theory, practice and person.

Korthagen, 2017, p. 399

Changing practice is difficult. However, we need to move beyond recognising there is a difficulty to helping teachers overcome it – and in particular help them create an ecosystem for children’s autonomy that is equally coherent and mutually reinforcing as the one for teacher control. It may be that many aspects of the ecosystem of teacher control are taken for granted, and that they need to be more explicitly challenged – in other words, that it is not enough to provide strategies for children’s autonomy if we do not make visible the existing practices that conflict with these strategies. Making these practices visible – and beginning to question their place – may also help teachers make visible the beliefs that underpin them. However, this needs to be balanced with helping teachers experience self-efficacy (Aelterman et al., 2016; Bobis et al., 2016), not only in their teaching in general, but for teaching towards children’s autonomy specifically. Our research suggests that those teachers that experience the most friction might also suffer the most from lack of self-efficacy, attributing their difficulties to a lack of competence.

In the previous chapter, I showed how the teachers’ engagement with the programme and their change in practice was dependent on their school context as well as wider institutional pressures. In this chapter, I have considered how programmes such as the Stepping Stones programme may better support teachers’ inquiries and change in practice by facilitating greater meaning-making and dialogue, fostering a feeling of trust and self-efficacy, and helping teachers understand how their own beliefs and assumptions may impact on their practice, sometimes in conflicting ways. Promoting children’s autonomy in the classroom therefore depends on understanding (and embracing) the complexity of teacher learning. As Opfer and Pedder argue (2011), teacher learning is characterised by systems within systems, under the influence of meso- and macro-level contexts as well as various micro-level dynamics. This also means that teachers’ change in practice towards greater autonomy for children will also require systems that support it. Changing practice in general may be hard, but by the accounts of the teachers in the Stepping Stones programme, supporting children’s autonomy and agency in learning whilst balancing the demands of standards and learning objectives as well as the needs of 30 children in a small space is particularly so. What this research suggests above all, is that without autonomy and agency of their own, and without an atmosphere of trust and safety, teachers are bound to reduce

children's autonomy to small isolated pockets. The more pressure they experience, and the smaller these pockets will be. Worryingly, recent policy documents in England emphasise standards, memorisation and tightly sequenced learning (DfE, 2019, 2021a, 2021b; Ofsted, 2018, 2019a, 2019b), whilst many Early Years professionals have voiced concerns that the current educational discourse, in particular in the Early Years, overemphasises formal didactic teaching and may lead to developmentally inappropriate practice (Jarvis, 2018; TACTYC, 2017; Williams, 2021; Wood, 2019). In parallel, headteachers have reported that the new frameworks are leading to increased pressure, workload and an emphasis on prescribed methods (e.g. NAHT, 2020; Roberts, 2021). In such a context, there is a large risk that England is heading in the opposite direction to one which would foster children's autonomy in the classroom and their inner motivation to learn.

Part II - Capturing Inner Motivation: Investigating the Reliability and Variability of the Leuven Involvement Scale

Chapter 7 - Methodology to Part II (RQ 2)

7.1. Overview of Chapter 7

Part II is concerned with evaluating the Leuven Involvement Scale (LIS) as a potential instrument to be used in further iterations of the Stepping Stones project. I focused on its reliability, and also investigated the extent to which it was stable or variable between classrooms, children and from one moment to another, as well as factors associated with any variation. In this chapter, I describe the rationale for choosing the LIS as a potential instrument, reviewing other ways inner motivation may have been captured and measured, as well as the theoretical underpinning of the instrument. I also describe the research questions that arose around the validity and reliability of the instrument as well as methods used to answer these questions.

Whilst the realist stance taken in this thesis fits better with the focus on traditionally positivist concerns for reliability and validity of Part II, I also take a critical approach to the validity and suitability of the instrument that goes beyond simply considering its psychometric properties but also more broadly question its meaning and use. In particular in the discussion (Chapter 9), I therefore continue adopting a methodologically-constructivist position which supports the view that all knowledge – including quantitative knowledge – is partial and constructed.

7.2 Methodological Background to Part II

7.2.1 Measuring Inner Motivation?

One important aim of the Stepping Stones project is to increase children's inner motivation to learn, i.e. that the reasons children are engaging in learning come from within, rather than from pressure or coercion. However, this refers to a process that is not only complex and nuanced, but happening internally. Yet for a programme to be delivered at scale, there needs to be evidence that this aim is being met. We therefore need to make visible what happens inside young children's minds and bodies. Here we revisit the thorny debate around

epistemology and methodology I have previously discussed, around what counts as evidence and how programmes are best evaluated.

On the one hand, given the nature of what we are interested in, it could be argued that qualitative methods would be more truthful and higher in validity. In addition, it has often been argued that children have a right to express their opinions and experiences (Freeman, 1998); children's participation in decisions that affect them has been recognised as a human right through the UN Convention on the Rights of the Child (UNGA, 1989), and this holds for their participation in research (Powell & Smith, 2009). Research using participatory methods and children's voice in particular (e.g. Clark, 2005) provides a useful model for how children may be involved in the programme evaluation. For example, methods such as the mosaic approach would provide both verbal and non-verbal ways for children to express their ideas (Clark, 2001). However, such research would always be limited to a few children, as it would be time consuming and costly, not only in terms of data collection, but also in terms of building the trust and communication required to be given access to children's private worlds. Not only this, but qualitative research is not without its own limitations. Using child voice and participation tends to focus on experiences and meaning making – for example, children may use photography to share places associated with particular feelings, or role play may provide opportunities for children to share their thoughts and feelings about different situations (Clark, 2005). However, because the present research is about the relationship between teaching practices and children's inner feelings of motivation, this would require on the part of the children a high level of awareness of how their internal states connect with their experiences – for example, young children may not be able to reflect on what teachers do and how it affects their inner motivation versus their desire to please adults. This is not to say that the children are too young to understand what is happening to them (Mashford-Scott et al., 2012), but that they may not have the metacognitive skills to deal with such a level of abstraction and to express it to researchers, verbally or otherwise. In addition, eliciting and interpreting children's voices is fraught with difficulties (Lomax, 2012; Pascal & Bertram, 2009; Waller & Bitou, 2011). There is a power imbalance between adult and child, and this still holds in the research process (Clark, 2005; Einarsdottir, 2005). If done incorrectly, child participation can be tokenistic, and children may be trying to second guess what they think adults want to hear, in particular when such research takes place in an educational context (Clark, 2005; Mick, 2011). Interviews might lead children to be mono-syllabic (Tizard & Hughes, 1984), and not all children may equally participate, either because of children's shyness (Clark, 2005) or because of the dynamics *between* the children, leading to some

voices being silenced (Lomax, 2012). In addition, even participatory methods eventually require an adult to interpret what children have expressed (Waller & Bitou, 2011). This is difficult work, as adults bring with them an adult view of children's experiences which may be different from children's own perspective. Therefore, it has been argued that careful observations by adults who can tune into children's perspective, for example as teachers do in the context of pedagogical documentation, might provide a helpful 'bridge' between children's views and research (Clark, 2005; Waller & Bitou, 2011). In other words, the important work of *listening* to children (Clark, 2001) may not need to take the form of interviews or visual methods. Finally, participatory methods are not neutral – by engaging in the evaluation process, the children may also engage in activities, discussions and reflections that affect what happens in the classroom (Lipponen et al., 2016), which would not only make it difficult to disentangle the effects of the participatory method from those of the teacher development programme.

This is not to say that these difficulties cannot be overcome, or that such research would not be valuable – and indeed in section 6.3.3 I discussed the need to consider children's experiences of autonomy and agency from their own perspective. The aim here, in outlining the tensions and practical difficulties with qualitative research with young children, is not to dismiss it, but to show that it is neither straightforward nor a panacea. In addition, whilst qualitative work is needed, it may not be enough. I argued in Chapter 1 that in order to be influential, the project needs to provide the kind of evidence that policy makers will value (EIF, 2015; Wrigley, 2018). In the context of a large-scale project, a quantitative measure of inner motivation would ensure this outcome is given the weight it deserves. Despite the reductionist nature of trying to measure such a complex, dynamic process, there is a large precedent for using quantitative, psychometric tools to measure latent constructs in psychology. Indeed, most research on motivation and curiosity relies on them. If we adopt a methodologically-agnostic approach, using a quantitative tool may not provide us with a complete picture of inner motivation in children, but it will provide us with a snapshot that can be studied at scale and with statistical analyses.

Nonetheless, there are limitations to how well we can expect such a complex internal state such as inner motivation to be captured by a number – in other words, whether we can find a *valid* instrument to measure inner motivation quantitatively in young children. Questions of validity concern both conceptual aspects of instruments as well as technical aspects: an instrument needs to capture the construct it aims to measure (face validity and construct validity), and it needs to do so accurately and precisely (reliability) (Newton &

Shaw, 2016). A valid instrument is therefore one on which inferences can be made about a given construct and where measurements can be trusted to be representative of what they are intending to capture rather than being due to error, bias or chance (Krippendorff, 2011). In the case of inner motivation, however, the question of validity also concerns whether an instrument can discriminate between inner motivation and other forms of motivation that are the result of pressure or coercion. Therefore, the question remains as to whether we can find a measure of inner motivation that will not mask important subtleties in the kinds of motivations children experience, and which we can use reliably for research purposes. It is this issue which this chapter is concerned with.

Having argued the need for a measure of inner motivation, or related constructs, the rest of this methodological background is dedicated to a review of existing measures and the presentation of a potential instrument, the LIS (Laevers, 1994). I finish with remaining questions around this instrument and the need for further research, where I outline the research questions for this part of the thesis; these are centred around the general concern of whether the LIS is an appropriate and reliable instrument for the Stepping Stones project.

7.2.2 Review of Existing Measures of Inner Motivation and Related Constructs.

Criteria for Appropriate Measures of Inner Motivation. Research on inner motivation has primarily been carried out with older students and adults, and measures targeting related constructs in young children are scarce. In the early stages of project development, I reviewed a large number of instruments used with children aged 4-to-6 and targeting a number of constructs related to inner motivation, including curiosity, interest, attitudes and engagement (see Appendix J). The aim was to find an instrument we could use to evaluate the programme's impact on children's inner motivation. Any instrument we eventually use in future iterations of the Stepping Stones project needs to meet a number of criteria for it to be appropriate:

- As discussed earlier, the aim of this research is to prepare for a large-scale intervention evaluation; therefore the instrument not only needs to produce quantitative evidence, but it needs to be sufficiently time effective that it can be used at scale.
- The programme will be carried out with Reception and/or Year 1 classrooms, and therefore requires a measure that is informative about children aged 4 to 6.

- Of particular importance to the present project is also that measures allow us to discriminate between what the child is motivated to do autonomously versus what the child is motivated to do because they are seeking rewards or want to please or comply with rules (controlled motivation). Any measure that conflates the two will present difficulties in interpretation.
- Finally, evaluating the impact of classroom practices on inner motivation requires that the construct measured must be malleable, in that it can be influenced by teaching practices.

This latter point relates to the conceptualisation of constructs related to inner motivation as either states that are prone to change from one moment to the next (such as mood or arousal), or dispositions that are stable from moment to moment (such as traits and attitudes). For example, curiosity has been studied as momentary feelings (e.g. Litman, 2008) and behaviours such as question asking (e.g. Jirout & Klahr, 2020) as well as a deep-seated personality trait, for example a preference for uncertainty (Jirout & Klahr, 2012).

In terms of the aims of the Stepping Stones project, it may be difficult to shift personality traits which are considered to be relatively stable given the short timeframe; for example, state-curiosity tends to be considered a stable attribute of a person (Litman, 2005). However, even dispositions might be influenced by the context in which they develop. For example, interest is thought to develop from past experiences (Renninger & Hidi, 2011) and a student's goal orientations can be altered by teaching practices (Ciani et al., 2010). States, by contrast, are under the influence of contextual factors (Chaplin et al., 1988), though they may also be under the influence of dispositions. For example, curious people are more likely to experience satisfaction on days they also experience curiosity (Kashdan & Steger, 2007), which suggests that curiosity as a state is under the influence of both personality traits and may fluctuate from day to day depending on moment-to-moment factors. Therefore, considerations of our chosen instrument must involve understanding patterns in the stability or variability of the construct, and what may influence it. In summary, our chosen instrument must be scalable, appropriate for 4- to 6-year-olds, focused on inner motivation and tapping states rather than traits.

I now describe the measures reviewed, by type of instrument. However, what may appear to be a linear sequence using selection criteria was in fact a back-and-forth process, piloting different instruments, revisiting previous assumptions and coming back to measures

and constructs that had previously been dismissed. However, for the sake of clarity, these false starts are not explicitly described. The purpose of this review is to provide the context and rationale for our choice of instrument, the LIS, by describing different types of measures as well as their respective issues and limitations. A full list of reviewed measures is available in the appendix (Appendix J). This was not a systematic review – in other words, the list presented in the appendix is unlikely to be exhaustive, but it presents a good snapshot of the types of measures typically used in this age group.

Child Reports. Self-report questionnaires provide participants with an opportunity to voice their beliefs and are widely used in psychometric research. In addition, large amounts of data can be collected in a relatively short time. One study successfully used Personal Digital Assistants with elementary school students (Malmberg et al., 2013) to produce a series of snapshots of motivation at different points in time during the day. This method, called the Experience Sampling Method (Hektner et al., 2006), allows children to record in-the-moment experiences which may be difficult to recall. However, this method has not been used with young children, and more generally self-report measures for young children are rare. Indeed, there are a number of issues posed by the use of self-reports in young children.

First, young children are not yet able to read and therefore self-reports must be conducted as interviews with a researcher, making them much less time effective. Second, young children may have difficulty understanding items and the response formats, and this ability may be related to their language skills (Baroody & Diamond, 2013). One implication is that items and response formats need to be simplified and their number limited. Typically children are shown puppets, toys or pictures associated with agreement or disagreement with an item (“I like reading”, Baker & Scher, 2002), or with opposite items (“I think reading is boring”, Baker & Scher, 2002) and they are asked to pick which puppet, toy or picture they are most like. Finally, participants must be able to reflect on their internal states and truthfully report on them. Young children may not have developed the skills to be aware of such internal states, and may not have the ability to communicate them (Chambers & Johnston, 2002). Self-reports also require participants to be aware of and recall past feelings and behaviours, which may not always be accurate (Hogan et al., 1997).

Possibly as a result of difficulties with using self-reports with young children, there are very few such measures available compared to self-reports for high-school and college students. Two measures relating to motivation that are appropriate for children aged 4 to 6 are the Children’s Feelings at School measure (Valeski & Stipek, 2001) and the Children’s Interest

Measure (Baroody et al., 2006), but the latter is not publicly available. A number of other measures also target early reading interest and share similar characteristics (Baroody & Diamond, 2013), but as they focus on reading items, they do not relate to classroom activities in general.

One additional limitation of these two self-reports is that they only contain items relating to positive and desirable attributes, which poses risks of social desirability as children may want to appear as though they like school. The items are also positively worded and so a positive response on the scale always corresponds to a positive score. This may result in response acquiescence, where respondents agree with every item to avoid disagreeing with the researcher. Previous work suggested that social desirability is very high in Primary school children (Le Courtois, 2017), and it is likely to be even more so with younger children. During a pilot for this research, I trialled the Children's Feelings at School measure (Valeski & Stipek, 2001) and a new instrument I had created to avoid response acquiescence (see Appendix K for a full description of piloted instruments). However, neither instrument proved suitable. The children appeared to respond randomly to the new measure, and the majority scored at the high end of the Children's Feelings at School scale. Other research with 5-year-olds has also shown that their responses to scale items tend to be positively biased and this results in extreme scores (Chambers & Johnston, 2002; Roebers, 2017). As well as social desirability, children of this age may be 'overenthusiastic' about their judgements of school. This reported enthusiasm may be genuine but not be reflected in their observed level of motivation. This could explain the discrepancy between children's own reports of interest or motivation and observer ratings or adult questionnaires (Baroody & Diamond, 2013), which are based on children's behaviour and which do not suggest such a high level of motivation (Pascal et al., 1998; Stipek et al., 1995). In any case, it means the measure is not informative for our purposes, as the low variability does not allow us to discriminate between different children or classrooms, or provide scope for an increase over time (e.g. as a result of an intervention).

Teacher Reports. Adults can understand more sophisticated items, have better recall and are better able to reflect on past behaviour compared to children. Asking adults to report on a child therefore overcomes some of the major obstacles of asking children directly. Teachers have a more wholesome view of the child's behaviour compared to an outside observer and therefore may be able to make better judgements about what is typical of a child's behaviour (Baroody & Diamond, 2013). However, teachers only see children in a very specific setting, and this may affect their view of a child. One study found that the correlation between

teacher ratings of curiosity in children and their achievement scores was artificially high compared to other measures of curiosity, suggesting a confound (Alberti & Witryol, 1994). There is a high risk that teachers would conflate compliance and good behaviour with interest and intrinsic motivation. In addition, some researchers have voiced scepticism at teachers' ability to report about student's feelings and thinking, though less so for primary pupils (Boekaerts, 2016). Finally, asking teachers to rate children as part of a programme evaluation is also likely to introduce bias as the teacher would have a stake in the results.

Parent / Caregiver Reports. Parents know their children well and therefore are sometimes asked to complete questionnaires about their child. Existing measures ask parents about activities children do in the home and their frequency (Baroody & Diamond, 2012; Hume et al., 2015). One main advantage of asking about children's behaviour in the home environment is that children typically have more freedom to choose what activities they engage in, and therefore this may be more representative of their interests.

However social desirability presents an important challenge to parent questionnaires as they may feel their child's behaviour reflects well or badly on them, and this may colour their responses. Responses are also likely to be confounded by the influence of the home environment. Caregiver reports may also suffer from low response rates, in particular in areas with families of low socio-economic status where parents may be reluctant to engage with schools or in areas where parents may not speak English confidently. Parents whose children like school are more likely to be comfortable filling in such questionnaires and this creates response biases. Finally, in order to be appropriate for our programme evaluation, a questionnaire would need to make links between motivation in school and behaviours at home. We are not aware that any such measure currently exists.

Tasks. Tasks are used in psychology to test a particular ability or trait, or a response to a situation. For example, children's curiosity can be understood as a preference for the unknown and one task asks children to choose windows to open depending on different amounts of information given (Jirout & Klahr, 2012). Other studies have measured the amount of time children spend exploring a toy (Bonawitz, 2010) or children's persistence and assurance-seeking when faced with difficulties (e.g. Sawyer, 2017; Stipek et al., 1995).

Another measure is the free choice task, a common measure of intrinsic motivation in experimental studies (e.g. Lepper & Greene, 1975). Children are introduced to a task and then given a period of time where they can choose to continue engaging with it or not. Usually

motivation is measured through time spent on the activity, but the activity can also be coded for more sophisticated behaviours (e.g. Doctoroff et al., 2016).

The advantage of tasks is that they measure the actual behaviour of participants, rather than having to infer it from observations or asking participants to reflect. Their limitation is that a task is considered an indicator of a broader latent trait or to be impacted by experimental conditions. Neither of these situations are appropriate for a programme evaluation, as the programme is unlikely to affect a deep-seated individual characteristic like curiosity in the short term, nor would responses be subject to experimental conditions.

Observations. Observations relating to motivation typically measure time spent in an activity such as reading during a free choice period (Baroody & Diamond, 2013) or engagement as an indicator of motivation or interest (Baroody & Diamond, 2013; Downer et al., 2010; Ponitz et al., 2009; Powell et al., 2008; Walsh et al., 2006). The operationalisation of engagement ranges from on-task behaviour (Ponitz et al., 2009) to signs of excitement (Walsh et al., 2006). Observations have the advantage of being less subject to bias if researchers are blind to the condition of an intervention, as they are less likely to have a stake in giving children a high or low rating. However, measures of engagement that target on-task behaviour only capture a superficial form of engagement and do not distinguish between engagement that originates from compliance and engagement that is autonomously motivated. For example, Powell et al., (2008) define active engagement as children having focused attention, and with the child's behaviour in line with the teacher's instructions. In the current climate in schools, teachers are expected to motivate students using rewards and sanctions and maintain high levels of engagement (DfE, 2011). Therefore, teachers are under pressure to produce an atmosphere of engagement, even if this is achieved through external and introjected means. The implication for observations is that unless instruments measure the kind of engagement that comes from an internal desire to engage then they will not be good indicators of inner motivation.

One instrument that seeks to make such a distinction is the Leuven Involvement Scale (LIS, Laevers, 1994), which I present in the next section in more detail. This instrument is related to theories of flow and motivation (Pascal et al., 1998) and distinguishes deep engagement from superficial engagement. Therefore, it may provide the level of detail necessary to capture in-the-moment inner motivation and is a promising measure for our project. Whilst other measures similarly target different qualities of engagement, such as the Individualized Classroom Assessment Scoring System (inCLASS, Downer et al., 2010), they

are not freely available – for example, training for the inCLASS costs upwards of \$1000. Therefore, the LIS is a much more accessible measure of engagement quality.

Review Summary. Having reviewed a large number of quantitative measures that were appropriate for 4- to 6-year-olds and could be used at scale, and assessed whether they could provide a valid measure of inner motivation and could be impacted by the programme, we settled on the LIS as a potential instrument. However, there is a need to further investigate the instrument, not only because it relies on subjective judgements which could be unreliable, but also because it is a measure of engagement rather than directly capturing inner motivation. Before further developing the research questions, however, I describe the theoretical underpinnings of the instrument, which contribute to its validity as a measure of inner motivation.

7.2.3 The Leuven Involvement Scale (LIS)

The Concept of Involvement. The LIS was developed for Early Years settings (Pascal et al., 1998). It distinguishes deep involvement from superficial activity, and therefore may provide the level of detail necessary to capture in-the-moment inner motivation. The concept of involvement emerged in action research in Early Years classrooms that was concerned with classroom processes rather than simply outcomes, and a desire to capture deep-level learning (Laevers, 1994; Laevers, 2003). It is defined as a dimension of human activity where individuals experience deep concentration along with “strong motivation, fascination and total implication” (Laevers, 2003, p. 15), akin to the state of flow described by Csikszentmihayli (1990). It is associated with energy and intensity.

The description of involvement therefore has strong ties to motivation, with intrinsic motives being given consideration as well as a lack of “calculation of possible benefits” (Laevers, 2003, p. 15). In other words, at its highest level, the involved individual is only acting out of intrinsic rather than instrumental reasons. Whilst making links with the concept of motivation, involvement is not anchored in any specific theory of motivation. The relationship between motivation and engagement is a thorny issue. Although engagement is a distinct construct from motivation (Boekaerts, 2016), motivation is often operationalised as engagement (e.g. Jang et al., 2010; Lepper & Greene, 1975). Not only this, but conceptually motivation theories such as Self-Determination Theory are often used to distinguish between different *kinds* of engagement, which is what the concept of involvement does. The original

description of the LIS (Laeyers, 1994) explicitly differentiates between: *intrinsic involvement*, which comes from the drive to explore and from interest; *emotional involvement*, which is seen as arousal and positive affect; and *functional involvement*, which refers to an instrumental reason for being deeply engaged in an activity. Intrinsic involvement is implied to be the type of involvement practitioners should be targeting. Involvement is also defined in terms of the challenge in the activity: the involved person is described as being in the “zone of proximal development” and “functioning at the very limits of his or her capabilities” (Laeyers, 2003, p. 15). Therefore the conceptualisation of involvement is relevant to our present interest in inner motivation.

Operationalising Involvement: The LIS. The LIS is the direct result of efforts to operationalise involvement in the context of Early Years classrooms in a Europe-wide project in the 1990s. It is a 5-point scale where engagement is rated from no activity to complete involvement (see the Methods section 7.4 for further details on the use of the instrument). Using the LIS has been described as “an act of empathy”: the observer “has to get into the experience of the child” (Laeyers, 2003, p. 16). Therefore, the essence of the LIS is that, drawing on observable signals, the observer infers what the child’s internal state of engagement is, and whether involvement is occurring. In this sense, it links to earlier concerns I discussed around children’s voice and the usefulness of observations by a caring adult open to the child’s perspective, and provides a method for doing this in a systematic, quantitative way. For example, Laeyers and Declercq (2018) have argued that by making children’s experiences visible in this way, the LIS can be a tool to support children’s rights and help them become equal partners in the shaping of educational experiences.

7.2.4 The Place of Involvement Within Engagement Theory

Before delving into more detail about the psychometric properties of the LIS, it is important to situate it within the broader context of research on engagement and different conceptualisations and operationalisations of this construct.

Different Conceptualisations of Engagement. Engagement is a broad construct that has been conceptualised in a wide range of ways, from in-the-moment interaction with an activity to general participation in school life (Boekaerts, 2016). Many studies rely on the tripartite typology of engagement first described by Fredricks and colleagues (Fredricks et

al., 2004), though some researchers have expanded on it (e.g. Jang et al., 2016). This typology identified three distinct dimensions of engagement: *behavioural engagement* refers to participation and attendance as well as conduct and on-task behaviour; *emotional engagement* refers to positive affect relating to different aspects of school (including teachers and peers), as well as attitudes, interest and values; and *cognitive engagement* relates to thoughtfulness and willingness to exert effort, as well as the use of strategies to pursue goals and solve problems (Fredricks et al., 2004). Studies have shown that these dimensions can be distinguished (Fredricks, Wang et al., 2016), but others have pointed out a large amount of overlap in the operationalisation of each dimension (Eccles, 2016).

In addition to these three dimensions, engagement has been studied either as a state or as long-term participation and attitudes – that is to say a stable characteristic of the student which has developed over time through previous experience, much as individual interests develop (e.g. Carter et al., 2012) or even as a personality trait (Hughes et al., 2008). There is variety in the operationalisation of engagement as a momentary experience as well: some studies rely on a single self-report for a specific lesson (e.g. Vasalampi et al., 2020), whilst others draw on methods that measure a student’s engagement at regular or random intervals, generating a number of data points per student (e.g. Heemskerk & Malmberg, 2020; Martin et al., 2015). Even then, some studies still consider engagement as a person-specific construct – that is to say, they aggregate the data for each student, and consider each student’s engagement in relation to other variables of interest such as learning (e.g. Downer et al., 2007; Salmela-Aro et al., 2016). However, increasingly studies in older students have shown that engagement fluctuates widely throughout any given day and that much more attention needs to be paid to this intra-individual variability (Martin et al., 2015).

Involvement as Deep Engagement and Links with Flow. Despite clear overlaps between engagement and the concept of involvement described in the LIS, it sits somewhat outside the research reviewed above. The LIS emerged out of research concerned with children’s experiential learning in Early Years classrooms and draws on a different body of research, though it makes explicit links to the theory of flow, a construct very close to that of involvement.

Concentration is considered a key sign of involvement (Laevers, 1994), and this is further described as attention that is so focused that a child would not notice or be distracted by events around them. Whilst flow is considered a ‘special case’ of engagement, and has recently been studied in the context of this body of research, it is usually considered distinct.

The term of flow is often used interchangeably with that of ‘absorption’, and it refers to a particular kind of high-quality engagement, where individuals lose themselves in an activity (e.g. Abuhamdeh & Csikszentmihalyi, 2012b).

Although descriptions of LIS do not specifically refer to Fredricks and colleagues’ tripartite typology of engagement (and the development of the LIS predated it), it is possible to see clear links between the two. The LIS refers to a number of characteristics of behavioural engagement, such as effort, persistence and activity. The presence of intense mental activity as well as problem solving and creativity as signs of involvement are aligned with conceptualisations of cognitive engagement. Involvement also presents an emotional aspect, with references to arousal and enjoyment, though on its own this is not considered enough to constitute true involvement. Therefore, all three dimensions of engagement are present in the conceptualisation of involvement, but the LIS is a global scale: children are given a single rating, rather than separate ones for different dimensions. On one hand, this means that issues with the ‘fuzziness’ in the operationalisation of the dimensions are avoided. On the other hand, it means that it is impossible to study the unique contribution of these dimensions. In research concerned with establishing mechanisms for engagement, these differences are paramount, since different dimensions may have different antecedents and learning outcomes. However, this may not be an important issue for establishing the impact of a programme because a single global outcome may be sufficient.

Before moving on, a final note must be made about the use of the words *engagement* and *involvement*. If involvement, like flow, is a state of intrinsically motivated and total absorption, then it only refers to the highest end of an engagement scale, but not the whole scale: a person is in flow or is not. By contrast, the LIS attempts to grade both the amount (engaged versus disengaged) and the quality of engagement, and therefore also measures more superficial forms of engagement, or even no engagement at all. In this sense, it resembles other measures of engagement as the construct is traditionally studied. For this reason, I use the term engagement throughout the remainder of this chapter, using the term *depth of engagement* to refer to the gradient along which the concept may be captured, both in amount and in intensity, and retaining the term *involvement* only for instances at the very end of the scale.

Engagement and Teaching Practices for Autonomy. Engagement is considered malleable and it is related to a wide range of classroom processes and practices (Fredricks, Filsecker, et al., 2016), which makes it well-suited as an outcome for our project. Of

particular interest to this research are those antecedents that are related to students' autonomy and self-determination. Research with older students has shown that when teachers increase the amount of autonomy support they provide (e.g. avoiding controlling language and extrinsic motivators, providing rationales and opportunities for authentic learning), this increases students' engagement (Reeve et al., 2004).

In addition, some aspects of instruction format are related to engagement. For example, instructional activities that are whole group teacher-led or directive are also related to less engagement compared to individual or peer-group activities, both in older Primary students in the US (Downer et al., 2007) and in England (Heemskerk & Malmberg, 2020) and in young children in the US (Powell, et al., 2008; Vitiello et al., 2012). By contrast, the closely related concept of boredom is predicted by lower feelings of control over the learning, as well as monotonous tasks lacking in mental stimulation and value (Pekrun et al., 2010).

Therefore it seems likely that engagement will be affected by the types of teaching practices that our programme is promoting. However, because previous research has been carried out with older children and/or has focused on off-task/on-task measures of engagement, there is still uncertainty as to whether the LIS, used with young children, would produce different results in an English context. Research on the LIS is an opportunity to confirm these findings with an instrument focused on depth of engagement, and in particular to investigate links with practices promoted by our programme.

7.3 Part II Research Questions (R.Q. 2)

7.3.1 Overarching Questions

As seen in the research reviewed above, there is a lack of good measures for investigating inner motivation and related constructs in young children. Whilst engagement is readily observed in classrooms, studies tend to focus on 'on task' behaviour. One exception is the LIS, which differentiates involvement (i.e. deep-level engagement) from more superficial forms of engagement (Pascal et al., 1998). Therefore, of all measures reviewed, the LIS seems to have the highest face validity for the age group of interest here, and data can be collected about children fairly quickly, which makes it usable at scale.

However, a number of questions remain before the LIS can be used in a project such as Stepping Stones. The general research question for Part II is therefore:

R.Q. 2 Is the LIS a suitable and valid instrument to be used in the Stepping Stones project?

First, we need to establish that the LIS can be used reliably. The LIS is limited by the ability of researchers to infer internal states and distinguish between them based solely on external signs. It is this property of the scale which also makes it powerful, as observing adults can notice small details and differentiate between closely linked emotions, and can attend to reasons children may have to engage in this activity, even when the children themselves do not yet have the capacity to do so. It therefore allows us a much more fine-grained understanding of what is going on for a child than self-reports would. However, the adult does not have direct access to the child's internal experience, and inferences are bound to be imperfect. This means that there will be some amount of error in the judgements made. How problematic this error is depends on how large it is, and can be investigated through reliability. Reliability, as defined by Krippendorff (2011, p. 1), is "the extent to which different methods, research results, or people arrive at the same interpretations or facts". In other words, we need to be able to "rely on known interpretations by others" (Krippendorff, 2011, p. 1) and this in turn relies on coders and readers of the research having a shared understanding of the phenomena being investigated. It is this shared understanding that leads to the reproducibility of the measurements, and hence leads to reliability (Krippendorff, 2011).

Inter-rater reliability of such a scale is therefore an important step in determining the validity of this instrument, as it determines how much error pollutes the dataset, and whether results are trustworthy. However, very little reliability information is publicly available about the LIS (Ishimine & Tayler, 2014). Originally, inter-rater reliability was reported using correlations based on 30 scores in real classroom situations (Spearman $r_{ho} = 0.9$, Maes and Nijsmans, 1988 IN Laevers, 1994), but the study was not published in the English language, making it difficult to scrutinise. Laevers (1994) also argues that "teachers possessing a fair amount of empathy are capable of handling the scale after a (half) day training" and a further study found that newly-trained raters could rate videos of children fairly accurately: 90% of ratings were within one level of the 'agreed' rating for the video (Pascal et al., 1998). However, percent agreements are generally considered poor indicators of inter-rater reliability, as they do not correct for chance agreements, and tend to give increasingly worse results as the number of categories increase (Krippendorff, 2011; Salkind, 2010). Previous studies using the LIS (e.g. Maes and Nijsmans, 1988; Pascal et al., 1998; Ulich & Mayr,

2002) did not provide traditionally-used statistics such as Cohen's Kappa or, better yet, Krippendorff's Alpha (Hayes & Krippendorff, 2007). Therefore the reliability of the LIS remains questionable.

Secondly, there is a lack of clarity as to the suitability of the LIS for the kind of research the Stepping Stones project plans to undertake. The LIS has been used both in preschool settings to evaluate childcare quality (Laevers, 2000; Pascal & Bertram, 1997) and in early childhood research (Barandiaran et al., 2015; Pascal et al., 1998; Ulich & Mayr, 2002), but it has not, to our knowledge, been used in pre/post programme evaluations. One important concern is whether changes in LIS ratings can be attributed to a change in teaching practice and therefore a genuine change in children's engagement, or whether engagement naturally fluctuates across different days. The latter would make interpretations of pre/post tests impossible. This depends on the stability of engagement over time. Whilst initial work on the LIS reported that levels of involvement within a setting tend to be stable (Laevers, 1994), there was little information given as to what this meant or how this assertion was arrived at. Laevers (1994) reports that teacher training can have an impact on the average engagement in the class, but no information is available as to how the study was conducted and the data analysed; in particular, it would be important to know whether the study used average scores at the class level, as has been done elsewhere (Pascal et al., 1998). In addition, whilst some research aggregates scores of involvement for each child (Pascal et al., 1998; Ulich & Mayr, 2002), other research suggests engagement fluctuates throughout the day (Heemskerk & Malmberg, 2020). Averaging, whether at the class or child level, may be misleading if it masks high variability. The lack of clarity as to where stability occurs and how this was investigated leaves a number of questions unanswered regarding the stability of the LIS.

Finally, the LIS was chosen because engagement is considered to be malleable and in particular can be influenced by the classroom environment (Jang et al., 2010; Laevers, 2004). The measure therefore has the potential to reveal the impact of teachers' practice on the children, and we were interested in investigating whether it was associated with classroom practices that might be promoted as part of our programme. In particular, some classroom practices may reflect the affordances children have for making decisions about their learning, such as activity setting. Activity setting can be thought of as the extent to which activities are self-directed or controlled by the teacher, from complete teacher control over the pace and content in whole class lessons, to child freedom and choice over stopping and starting activities in free flow; within this framework, small group teacher-led activities and

independent work can be thought of as requiring intermediate amounts of self-directedness. Therefore, activity setting provides an easily observable way to categorise the extent to which children's activities are self-directed. Whilst self-directedness is not the same as autonomy (as argued in Chapter 5), children in free-flow settings are more likely to engage in play-based activities (Walsh et al., 2006) and to have opportunities to choose tasks of interest to them, and therefore we argue they would have more opportunities to experience autonomy. By contrast, children in whole class settings all receive the same instruction which may not support their feeling of autonomy or their inner motivation. For example, Renninger and colleagues (2018) have found that whole class situations have fewer triggers for interest than spontaneous or individual activities.

As well as theoretical support from research on autonomy described in previous chapters, research using the LIS highlights the importance of opportunities for exploration and initiative (Laevers, 2003). In addition, a number of studies show an association between activity setting and engagement. Previous research has found that whole group teaching is on average the least effective for children's engagement and motivation, (e.g. Godwin et al., 2016; Powell et al., 2008; Vitiello et al., 2012) but most studies relied on traditional measures of on-task/off-task engagement² and were carried out in a different context, the US, where the form these activity settings take might be different to English classrooms. They are also limited to a small number of correlational studies. Using a measure of engagement that explicitly distinguished between passive and internally motivated engagement, we wanted to further investigate the link between activity settings and engagement in our context of interest, i.e. early Primary classrooms in England.

7.3.2 Pilot Study Summary

In order to begin investigating the validity and reliability of the LIS, I conducted a pilot study investigating inter-rater reliability and stability of the LIS. Here I present a brief summary of methods and findings.

Pilot Methods. 48 children aged 4-6 from six classrooms in four schools took part in the pilot. Researchers observed each child in their classrooms for two minutes, then rated the

² The measure used by Powell and colleagues has a category called 'active engagement', but it contains little that differentiates it from on-task behaviour, and does not include the sorts of details captured by the constructs of involvement or flow, and makes no reference to motivation. The study by Vitiello and colleagues uses the inCLASS (described previously), which makes references to intensity and initiative.

child's engagement in the activity using the LIS. For each observation point, the researcher also made notes about the activity the child was involved in. Children were observed on three occasions each day, and 73% of these children were also observed on a second day, resulting in a total of six observations. We coded 252 events in total, of which 122 events were simultaneously coded by two researchers.

Notes about the activity were coded into: whole class teaching, independent work, free play and other. Observation procedure and analytical approaches were identical to those described for the main study in the methods section and in Chapter 8. One exception is that stability was calculated using the ICC (Intra-Class Correlation coefficient), using only the dataset spanning two days.

Pilot Findings. Further details of analyses can be found in Appendix L. Inter-rater reliability was high (K-alpha= .81, CI= 0.72 - 0.87), which suggests that the observations made with the LIS were reliable. Reliability increased over time, likely because we compared our ratings at regular intervals, and therefore our common understanding improved over time.

Stability, however, was poor: there was no correlation between a child's engagement at any given point in time. Multilevel models confirmed these findings, with no variability attributable to the child. The school and the class uniquely accounted for a small amount of variation (3% and 5.5% respectively in the empty model; 3% and 0% in the final model), with most variation occurring between observations. This suggests that it is the individual moment, rather than structural aspects of the classroom or characteristics of the child, which are responsible for differences in engagement. However, this was based on only six ratings per child. Whilst previous studies have also used six observations split over two sessions (Bertram and Pascal, 1997), this may have masked more stable patterns over time and therefore increasing the number of observations per child might provide different results.

We also investigated what aspects of the context of each individual moment might affect engagement by investigating the association between activity setting and LIS ratings. Compared with being in a whole class teacher-led situation, being in a free-flow setting was associated with a significant 1-point increase on the rating scale ($p < 0.001$), whilst there was a significant 0.5-point increase on the rating scale when children were in independent activities ($p = 0.004$). This suggests that activity settings, and the opportunities they afford for children to lead their learning, are an important aspect of the individual moment which are associated with engagement.

7.3.3 Final Research Questions for Part II (R.Q. 2)

The aim of the main study in Part II was to further investigate and confirm the findings of the pilot with a larger sample size. In particular, by having more observations per child, we wanted to ‘stress test’ the finding that variation in LIS ratings were largely due to individual moments, and not at all associated with differences between children. Research questions are split into three main questions, each with specific sub-questions.

R.Q. 2.1 Is the LIS a reliable measure of engagement?

R.Q. 2.1.1 Can we achieve good inter-rater reliability using the LIS?

Whilst the inter-rater reliability in the pilot was good, it was based on a small number of ratings and these were not always independent because we regularly ‘checked in’ with each other’s ratings. Therefore, we investigated inter-rater reliability again in the main study, first using live coding then recording videos and using video coding.

R.Q. 2.1.2 Can we achieve good intra-rater reliability using the LIS?

It may not always be possible to have two or more observers, and so we may need to rely on a single observer’s interpretation. It is also important that “observation drift” (Pascal et al., 1998) does not occur, which would lead to differences in the meaning of the data depending on when they were collected. Therefore, we also asked whether observations were reliable across time for the same observer, using videos coded twice at different points in time.

R.Q. 2.2 Is the LIS stable over time?

R.Q. 2.2.1 Is there stability between observations of individual children?

The pilot suggested that there was little stability between observations of individual children. We therefore hypothesised that there would be a large amount of variation in engagement between observations.

R.Q. 2.2.2 Is there stability at the level of the classroom between engagement on different days?

In the pilot, the finding that there was no stability in engagement was based on a small number of observations, and so it may be that the variation in engagement forms a stable

pattern over time. Specifically, we hypothesised that the day on which observations are made would not have a significant impact on children's engagement.

R.Q. 2.3 What accounts for variation in engagement?

R.Q. 2.3.1 How much variation in engagement occurs between children and between classrooms, compared to within children and within classrooms?

Whilst in the LIS engagement is studied as a state, research that aggregates ratings at the child level (e.g. Pascal et al., 1998; Ulich & Mayr, 2002) assumes, at least implicitly, that engagement is partly a characteristic of the child. This links to the questions of stability addressed above and therefore requires us to determine how much variation occurs between classrooms, between children and from moment to moment. The pilot suggested that there was no variation at all between children and very little between classrooms, but in the main study we greatly increased the number of observations per child (from 3-6 observations per child in the pilot, to 17-30 observations in the main study) in order to capture any stability in engagement within children, if it exists. Therefore, we expected that a small amount variation might occur between classrooms and between children. This means that children within a particular classroom will be slightly more similarly engaged compared to children in other classrooms, and that observations taken for a single child will be slightly more similar than between children (even within the same class).

R.Q. 2.3.2 What contextual factors are associated with engagement?

Whilst some stability in the features of classrooms is expected, many aspects of the context are likely to vary over time, and this may explain fluctuation in engagement throughout the day. Understanding this fluctuation is important because we hope to raise engagement through our programme and because it may contribute to discussions around teacher practices that support children's engagement. We focused on activity setting, that is to say whether children were in a teacher-led whole class setting, in a teacher-led small group setting, doing independent work (collaboratively or individually) or in 'free flow' (i.e. freely chosen activities). In the pilot, we found that activity setting was indeed associated with engagement and we wanted to further investigate this relationship with a larger sample size. Therefore, we predicted that free flow would be associated with higher engagement as it provides the greatest opportunities for self-direction.

R.Q. 2.3.3 What child characteristics are associated with engagement?

Given the increased sample size and the conceptualisation of engagement as a child-level construct in a large number of studies, we hypothesised that there may be some differences in engagement between children, and wanted to investigate what might be responsible for these differences. In particular, given the association between engagement and self-regulation, we wondered whether aspects of children's regulatory behaviours at school may explain child-level differences in engagement. We therefore included two child-level measures of child behavioural characteristics, the Strengths and Difficulties Questionnaire (SDQ) and the Child Behaviour Questionnaire (CBQ), described further in the methods section. These include sub-scales on children's ability to focus, to engage positively with peers and to deal with difficulties, which could be related to engagement.

7.4 Methods for Part II

7.4.1 Description of the LIS

The LIS (Laevers, 1994) is an observation instrument rating children's engagement on a scale of 1 (no engagement: the child shows hardly any activity) to 5 (intense engagement: the child is continuously engaged in the activity and completely absorbed by it). Children's engagement is rated based on nine signals: concentration, energy, creativity, facial expression, persistence, precision, reaction time, language and satisfaction. Observers do not rate children on individual signals, but use these to focus their attention. Each signal is described according to key behaviours. For example, concentration is described by the statements, 'The attention of the child is directed toward the activity. Nothing can distract the child from his/her deep concentration.' These signals need to be interpreted by the observer, who needs to decide on their relative importance for giving the final rating depending on the context and what they know of the child. For example, whilst the child's focus is considered essential for a high rating, other signals such as pleasure might not always be visible, so a child could be rated highly even if they are not visibly expressing enjoyment if they are otherwise displaying personal commitment. Similarly, verbalisations are taken into account in the signal of 'language', but not all situations will lend themselves to children using language to express involvement, and this should not lead to the rating being downgraded. This requires the observer to exercise judgement in deciding to which signals they ought to give weight in any given situation, and to infer what they might reveal about the child's involvement. It also means that the same signal may be interpreted differently depending on

the situation. To help with this, each level on the rating scale corresponds to an overall description and indicator behaviours relating to the signals. For example, a rating of 3 (which corresponds to moderate involvement, or passive engagement) is described as: ‘the child is busy the whole time but without real concentration’. Different authors have described each level in slightly different ways, so these different versions were compiled into a single instrument for the present research (Appendix M). Additional comments were added during the pilot to help with clarity, in particular in settings where children did not choose their activities.

7.4.2 Participants

65 children (43% female) from 13 classes across 8 schools in the East of England were observed in their regular classrooms (Table 2). This excludes one child who was removed from the dataset as they were taken ill on the first day and not enough observations were made. Children were in Reception (62%) and Year 1 (38%) and aged 4 years 3 months to 6 years 8 months (average: 5 ½ years). Within each classroom, five children were randomly selected from the list of participating children using a random number generator.

Table 2 *Sample Sizes for the LIS Study*

Data collection dates	Live coding	Video coding
	April 2019 – February 2020	November 2019 – March 2020
Number of schools	8	4
Number of classrooms	13	7
Number of children (%Reception/%Y1)	65 (62/38)	34 (70/30)
Total number of observations	1669	112 (+ 38 videos for training)
Number of observations per child	17-30	2-7
Number of observations on which reliability was calculated	40	112 for inter-rater reliability 42 for intra-rater reliability
Number of observation days per class	3	1

Note. The children in the video coding study were also part of the main study but video-recorded on separate occasions, on an additional day.

34 of these children (56% female; average age 5 years 3 months) in 7 of the classrooms were also video recorded in an additional session. These were the last 7 classrooms to be visited, rather than the result of a selection process.

Sampling and School Information. I used purposive sampling to recruit schools (Table 3). Schools were targeted to represent a wide range of teaching practices and ensure some variation between schools.

Table 3 *School Characteristics in the LIS Study*

School	State/ private	Percentage bands ^a of children eligible for deprivation premium (schoolwide, 2019)	Classes	Year group
2A	state	0-10	Class 2.1	Reception
			Class 2.2	Year 1
			Class 2.3	Year 1
2B	state	20-30	Class 2.4	Year 1
			Class 2.5	Reception
2C	state	0-10	Class 2.6	Reception
2D	private	N/A	Class 2.7	Year 1
			Class 2.8	Reception
2E	private	N/A	Class 2.9	Reception
2F	state	10-20	Class 2.10	Year 1
			Class 2.11	Reception
G	state	0-10	Class 2.12	Reception
			Class 2.13	Reception
H	state	10-20		

^a The percentage of children eligible for ‘Pupil Premium’ is generally considered in England to be an indicator of socio-economic deprivation. It is represented here in the form of bands to preserve schools’ anonymity. The national average is 15%. ^b Not all classes in a given year group were visited.

Schools were contacted by email or through existing contacts. In addition to feasibility issues, difficulties in recruiting and accessing schools meant random targeting of schools would not have resulted in a random or representative sample. Information about ethnic background or

socio-economic status for individual children was not collected because the child-level measures were deemed to be more directly associated with variation in engagement.

Ethics and Consent. Teachers interested in the study received a Memorandum of Understanding (Appendix N), outlining what their participation would require and the responsibilities of the researcher; this was co-signed by the researcher and the teacher if they agreed to take part. Teachers did not receive benefits in kind other than a small token of appreciation, but they were able to see anonymised analysed data for self-reflection purposes. Children in participating classrooms received a letter for their parents outlining what the study would involve; for some children, these were letters requiring opting out if parents did not wish their child to be considered for the study, as observations are routinely carried out in classrooms and would not disturb the children's learning, and schools acted as 'gatekeepers'. Children who were also video recorded (for video coding) in addition to classroom observations needed to 'opt in' in order to be considered for participation (Appendix O). Whilst parents gave consent on behalf of their child and children were not asked for verbal consent, observations were interrupted if children showed verbal or non-verbal signs of being uncomfortable with the observations.

7.4.3 Observation and Rating Procedure

Overall Procedure. I observed each child in their classroom for two minutes, then rated the child's engagement in the activity using the LIS. A round of observation continued until all participating children had been observed, then a new round of observations began. Children were only observed during classroom activity; this excluded transitions and other times where we had found the concept of involvement was difficult to apply during the pilot (e.g. lining up, tidy up time, snack time). In the video coding, researchers coded two-minute segments of four-minute videos, with approximately the first and last minute used for additional context.

Live Coding. I carried out rounds of observation throughout the day for three days in each classroom. Between 6 and 10 observation rounds were carried out each day, resulting in a total of 21 to 30 observations per child, except for one child who was absent on the last day of observation and for whom only 17 observations were made. One further child was excluded from the study because she was taken ill on the first day and only a few observations were made, and two additional observations were excluded because they did not meet the

requirements for inclusion; in one case the activity was interrupted by the teacher (therefore not a single context category) and the other was a tidy-up activity. There remained 1669 coded events and 40 of these events (2% of total) were simultaneously live coded by another trained researcher.

Video Coding. Videos were taken throughout the day in the children's normal classroom activities, excluding transition points. Videos shorter than 4 minutes because of teacher interruptions (e.g. asking children to stop an activity) were discarded. 38 videos were used for training purposes and 112 videos were used for inter-rater reliability with another researcher. I also recoded 42 videos after a 12-week interval to test intra-rater consistency. Coding started at 1 minute (unless otherwise indicated on the record sheet) and stopped after 2 minutes, but researchers viewed the whole clip for context. I determined the exact start-time for coding based on the start of a new activity and shared it in the recording template so all researchers would observe the same segment. I also provided context notes which would not have been evident from watching the videos without knowledge of the classroom, for example whether children had chosen the activity.

7.4.4 Training

All researchers (including myself) were trained with videos from UK classrooms and a manual developed by the University of Leuven and Early Excellence, a UK-based Early Years consultancy (Laevers et al., 2010). In addition to these training videos, for live coding the additional researcher was also trained during one day of live coding in a local classroom. For video coding, a different researcher was trained using the above videos (Laevers et al., 2010) as well as additional videos (Laevers et al., 2011), both from the University of Leuven and Early Excellence, and 38 videos collected as part of the main study. For both live coding and video coding, training took place in two phases. The first phase used the DVDs and encouraged the researcher to learn how to use the instrument and check their understanding against the manual. The second phase was focused on practice, using either live coding in a classroom or further videos, with rounds of coding followed by a debrief to compare our interpretations. Reliability during the practice phase was then calculated.

7.4.5 Context Coding

Because of the focus of our research, we also hypothesised that activity setting might be responsible for some of the variation in involvement. Notes describing the event and the context were taken during the observations and these were later recoded into a categorical variable to capture the setting of the activity. Setting was coded as being either: whole class or large group led by the teacher (hereafter called ‘whole class’); teacher-led small group or one-to-one (hereafter called ‘teacher-led small group’); independent individual or small group work set by an adult but without their direct guidance (hereafter ‘independent’); and free flow or carousel (hereafter called ‘free flow’), in which children could freely choose between activities within a particular context (e.g. indoor play, break time, forest school). The activity setting code distinguishes between differences in opportunities for children to self-direct in their learning (see Appendix P for further description).

7.4.6 Child-Level Variables

Teachers were asked for children’s gender, as well as date of birth which was converted into age in months at the time of the visit. While these child factors were not directly of interest in the current study, previous research has suggested that age and gender may impact on engagement levels (Lietaert et al., 2015).

Teachers also completed the Strength and Difficulties Questionnaire (SDQ, Goodman, 1997) and the very short teacher form of the Children Behaviour Questionnaire (CBQ-VSF-T, Putnam & Rothbart, 2006; Allan et al., 2013) to assess the association between engagement and individual differences in classroom behaviour indicative of latent psychological attributes. The SDQ is a brief behavioural screening for children aged 3-16 years old, and captures 25 attributes, both negative and positive, across five scales: 1) emotional difficulties, 2) conduct problems, 3) hyperactivity-inattention, 4) peer problems, and 5) prosocial behaviour. The CBQ-VSF-T captures individual differences in 3-to-7-year-olds’ reactivity, affect, and self-regulation through 15 items across three subscales: 1) surgency, 2) negative affect, and 3) effortful control. The SDQ and CBD assess a range of individual differences in children that are likely to impact on classroom engagement. The validity and reliability of these instruments have also been tested widely and in a range of contexts (e.g. Dahlberg et al., 2019; Hall et al., 2019; de la Osa et al., 2014).

7.4.7 Sample Size Determination

In all three studies, data were nested: observations were made of children, which were part of classrooms, within schools. Depending on the research question and the analytical approach, this is more or less relevant and problematic to sample sizes.

Reliability. Inter-rater reliability calculations were made using a statistic (Krippendorff's alpha, or K-alpha, Hayes & Krippendorff, 2007) for which there are no sample size or power calculations available. To determine the sample size, I used estimates from Bonett and Wright (2000) for Spearman's rank correlations, which are identical to the K-alpha in its two-rater ordinal form (without ties in ranks), as was the case here. Because sample size depends on the expected correlation value, type I error rate and specified confidence intervals, a wide range of sample sizes could apply. Sample sizes well above 100 were unfeasible in the present research as it was limited by the number of observations and videos. Instead, I aimed for 50-100 datapoints, which corresponds to CI=0.3 for a correlation coefficient of 0.6 or 0.7, and CI= 0.2 for a correlation coefficient of 0.8 (assuming $\alpha=0.05$) – these correspond to K-alpha values we could expect. Bootstrapping was then used to establish confidence intervals based on the actual data and K-alpha.

Stability. Stability was investigated as part of other investigations around variability, and therefore considerations of sample size are best understood in that context (see section below). However, care was taken to observe classrooms on different days to establish stability over different days, which affected decisions around the number of children and classrooms that could be observed.

Variability and Contextual Factors. As described in further details in Chapter 8, this area of investigation relied on multilevel modelling, where the sample size at higher levels (here, the child and the class) are paramount. The issue is that obtaining large sample sizes at the higher levels are difficult and costly, whilst small sample sizes result in analyses that can produce unreliable estimates (Hox et al., 2010). Studies have shown that multilevel models tend to produce unbiased regression coefficients even at low group sizes, but that variance estimates are biased in groups considerably smaller than 100 (Maas & Hox, 2005). However, some studies have found good estimates for group sizes as small as 12 (Browne & Draper, 2000) when using modelling that can account for smaller groups, whilst others have

found highly biased variance estimates unless groups are at least as big as 30 or even 50 (Maas & Hox, 2004).

Sample size at the observation level was easily large, as 30-50 observations can be made on a single day, and data collection was spread over a number of days to allow us to study stability and variability over time (as described above). However, decisions had to be made in terms of how to split these observations between children and classes, and with limited resources this entailed trade-offs.

Snijders (2005) argues that cluster size is not important for the power of analyses in terms of regression coefficients, but it does affect the power for testing random slope variances at the higher levels. In our case, the number of observations per child (i.e. the cluster size at observation level) was important to distinguish between-children from within-child variation, as well as for the random slope models at child level. By increasing the number of observations for each child and therefore for each classroom, the intention was to ‘stress-test’ the results of the pilot (i.e. that no variation was explained by the child level or the class level; see Appendix L for further details). However, time and resource constraints (as well as difficulties in recruitment) meant that this involved limiting the number of children per classroom, and limiting the total number of classrooms, therefore having a smaller sample size at higher levels.

Initial power analyses (Appendix Q) suggested that 5 children in 10-35 classrooms would result in appropriate power, using expected effect sizes based on the pilot and other published work (Vitiello et al., 2012), for observation-level predictor variables only. I aimed for 20 classrooms because it was the maximum number of classrooms I could feasibly visit in the remaining time of this PhD, and because this corresponded to 100 children which is the sample size at which samples are no longer considered ‘small’ and therefore problematic (Maas & Hox, 2005; McNeish, 2017). However, the start of the COVID-19 pandemic in 2020 curtailed data collection at 13 classrooms and 65 children. The sample size at child-level was still acceptable, though on the low-end, with 50 being the accepted cut-off point (Maas & Hox, 2005; Hox, 2010; McNeish, 2017). However, the sample size at class level was even smaller than anticipated. Because I was not planning on including any classroom explanatory variable, this did not affect the power of the fixed intercepts in the analyses, but it did affect the random slope models at class level as well as the reliability of the variance at class level, as per the criticisms of small samples described above.

In this regard, the present research is in good company – small samples at higher level are fairly common, with one study reporting that as many as 20% of published multilevel

models fell under this umbrella (McNeish, 2017). As a result, the difficulties of increasing sample size at higher levels are well-recognised, and the impact this has on model statistics well-researched (McNeish, 2017). I included available methods for dealing with small samples in the analyses and I discuss the limitations of the small sample size at the classroom level in Chapters 8 and 9.

Chapter 8 - Part II Analyses and Results

8.1 Reliability of the LIS: Analytical Approach and Results for R.Q. 2.1

8.1.1 Analytical Approach

Inter-rater reliability was calculated using Krippendorff's alpha (K-alpha; Hayes & Krippendorff, 2007), because it is appropriate for ordinal data and because software exists to calculate it easily. In addition, K-alpha has been shown to be a better measure of reliability than other statistics such as Cohen's Kappa (Hayes & Krippendorff, 2007; Salkind, 2010; Zapf et al., 2016) because it calculates disagreements rather than chance-corrected percent-agreements. This avoids the limitations of other statistics, which might be limited to nominal data, or result in decreases in reliability with more raters or more categories (Hayes & Krippendorff, 2007). It can also be used with missing data (Zapf et al., 2016). K-alpha assumes coders follow the same coding instructions and work independently of each other (Salkind, 2010).

I used the online ReCal calculator for quick checks (Freelon, 2013), and redid the tests including bootstrapping in R (R Core Team, 2020) using code written by Zapf et al. (2016) in order to produce the confidence intervals. Type-one error was specified as 5%, and the number of bootstrap samples was 1000.

8.1.2 Results on Reliability

Live Coding. Despite achieving good inter-rater reliability during the pilot, initial live coding in classrooms resulted in poor inter-rater reliability, with K-alpha = 0.54 (CI= 0.29 – 0.71), which is underneath the 0.6 threshold considered for tentative results, and well below the 0.8 threshold for good reliability. However, even at the end of training reliability had been at the lower end of acceptability, with a large confidence interval (K-alpha = 0.66, CI = 0.38 – 0.81). Unfortunately, because of time pressures we did not notice this in time to improve reliability. Poor reliability seems to have been caused by differences in interpretation in what had happened in the observation period. For example, a child who was waiting compliantly could be interpreted as highly engaged in 'waiting', or conversely it could be argued that there was no activity to be engaged in. Similarly, a child looking through a picture book with focus and care, but without actually reading the words, could be interpreted as only superficially engaged in reading, or highly engaged in a pre-literacy activity. As a

result, only data I had collected myself were used in the following analyses, and my consistency was checked using videos to ensure internal consistency in the models.

Video Coding. To further establish inter-rater reliability, I carried out video coding with a different researcher using videos filmed in the same classrooms as the live coding but on different days. During training, reliability was good (K-alpha = 0.76, CI= 0.60 – 0.86). Inter-rater reliability calculated on 112 independent ratings was 0.68 (CI= 0.58 - 0.76), which is acceptable.

For comparison purposes with regards to other studies, percentage agreement within one rating of each other (the traditionally reported measure of agreement in studies using the LIS) was 83% for live coding and 96% for video coding, with 45% total agreement in video coding. In conclusion, it is possible to achieve acceptable reliability with the LIS, as long as a common understanding is reached regarding how to interpret different events.

I also established intra-rater reliability using these videos to ensure the consistency of the ratings used in further analyses. Based on 42 videos I coded on two occasions at a twelve-week interval, my intra-rater reliability was good (K-alpha = 0.81; CI = 0.65 - 0.89). Agreement within one rating was 100%, and total agreement was 61%.

8.2 Stability of the LIS: Analytical Approach and Results for R.Q. 2.2.

8.2.1 Analytical Approach

I used plots to visually examine variability at the child level and at the class level. The high variability evident in the dataset meant that using correlations of averages (e.g. child daily engagement, or class-level daily engagement) would mask this variation and was therefore not appropriate for establishing stability. The pilot had indicated that there was little stability at the level of individual observations (i.e. from one observation to the next), as seen in the low ICC. Here, data were unbalanced (there were different numbers of observations for different children), so I could not calculate the ICC. In addition, our concern was whether stability emerges at the classroom level from one day to another. Traditional measures of test-retest reliability or stability, such as the ICC, are not suitable, as they are restricted to correlations between single observations.

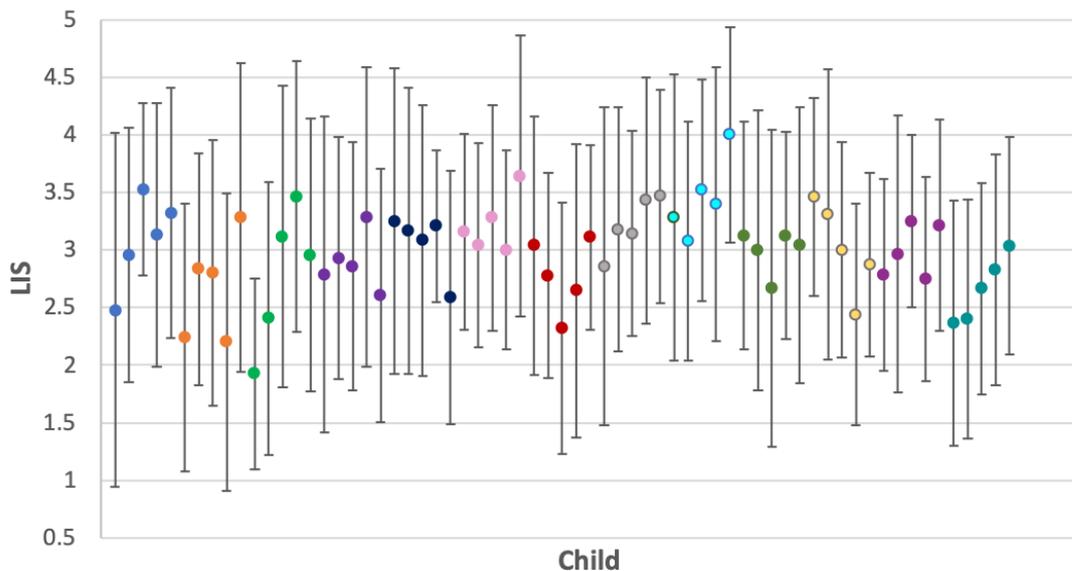
Instead, I used cross-classified multilevel modelling to investigate variation in engagement on different days, with days crossed with the other levels (child > class > school).

This accounted for the nested nature of the data and therefore allowed us to tease apart the within-day variation in different classrooms from the between-day variation in order to examine whether there were important differences in engagement depending on the day. As each ‘day’ does not represent a variable that is the same across classrooms (day 1 for class 1 is not the same day as day 1 for class 2), they were recoded so that each classroom had a different set of days (i.e. days 1, 2 and 3 for class 1; days 4, 5 and 6 for class 2; etc.). The general procedure for multilevel analyses is described further in a later section.

8.2.2 Results on the Stability of the LIS

Visually, as with the pilot there was no stability between observations because of high variability (Figure 9). This variability, and how much of it can be explained by different factors, are further explored later.

Figure 9 Mean LIS and Standard Deviation for Each Child

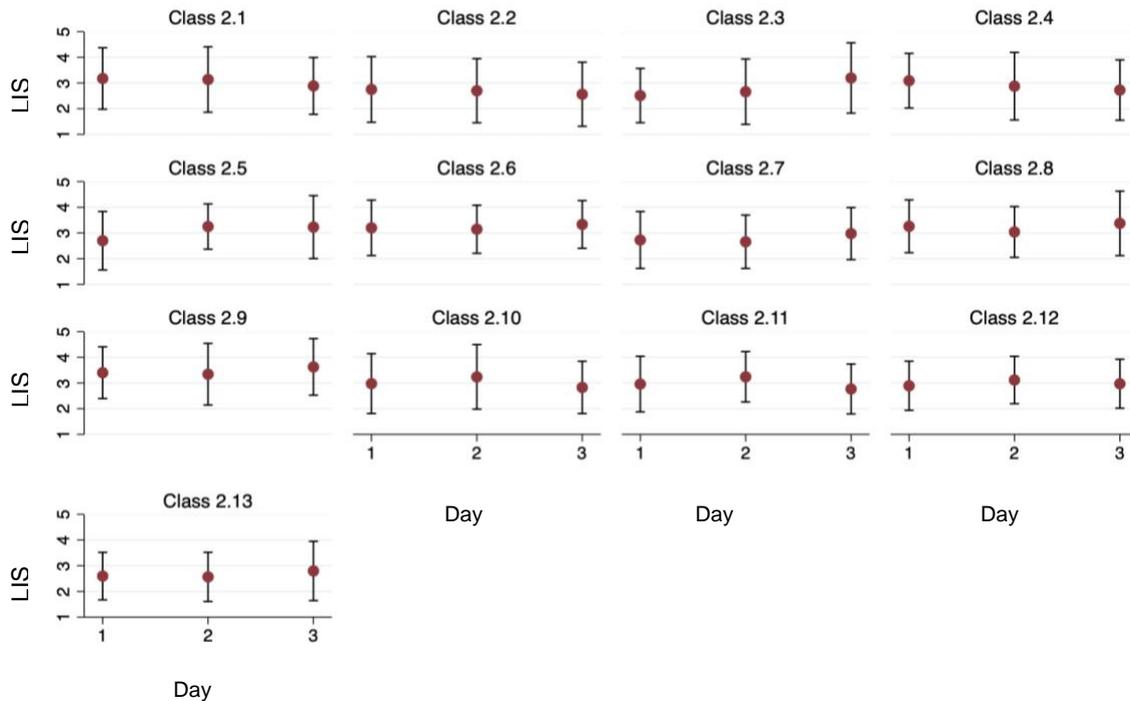


Note. Colours represent different classrooms.

Cross-classified models showed the day did not account for any amount of variation in engagement (random effects <0.01). In other words, overall the day on which we came to a classroom did not explain any of the variation in engagement. However, this does not mean that individual days might not have been associated with different amounts of engagement – that is to say, it’s still possible that in any given class, engagement would be higher on some

days than on other days. What our results suggest is rather that, overall, differences between days are negligible compared to the within-day variation (Figure 10).

Figure 10 Mean LIS and Standard Deviation by Visit Day and by Class



8.3 Variability and Factors Associated With the LIS: Analytical Approach and Results for R.Q. 2.3

8.3.1 Analytical Approach

Multilevel Models. Multilevel modelling was used to describe the amount of variation in engagement at the child level and classroom level compared to variation that can be predicted by the observation level and context variables (i.e. activity setting), taking into account the nested nature of the data. Multilevel models were necessary because grouped data, such as was the case here, violate the assumption of independence of all observations, even when the analysis includes only variables at the lowest level (Maas & Hox, 2005; Snijders & Bosker, 2012), in this case the observation level. Ignoring clustering leads to incorrect standard errors and therefore incorrect significance tests.

Models were run sequentially in Stata v16 (StataCorp, 2019) using a “bottom up” approach (Hox et al., 2010), which is to say that the simplest models are fitted first and then increasingly complex models are fitted, checking the improvement in model fit with a

difference in deviance test at each step. In our case, this means I first fitted empty models, then random intercept models, then random slope models, first at the child level, then the class level, then both (as per Hox et al., 2010). At each point, a likelihood ratio test was carried out to determine which was the best model (the difference in deviance test). Similarly, level 1 explanatory variables (activity setting) were included first, then level 2 variables (child characteristics). Restricted maximum likelihood with Kenwood-Roger correction can be used to account for the small sample size at child and class levels (McNeish, 2017). However, the correction means it is no longer possible to use likelihood ratio tests as the restricted methods means fixed effects are no longer comparable across models. Alternative tests to compare models such as Akaike's Information Criterion (AIC) are not considered as reliable as the likelihood ratio test because they penalise models for complexity, i.e. having more variables in the model (Hox et al., 2010). Therefore the models were initially run with maximum likelihood methods, and once the final model was selected it was run again with restricted maximum likelihood with Kenwood-Roger correction.

Random effects were not allowed to covary. Although it is possible they would (e.g. a child with better executive function might be better at staying engaged in a range of situations, and a class teacher who better supports small group work might also be better at supporting whole class work), it made the models too complex for the present dataset, and initial analyses at child level did not support the hypothesis that random effects covaried. After each model, the ICC was calculated to show how much variation occurs at each level. The ICC is not a measure of the model's explained variance (Roberts et al., 2010), but rather shows the extent to which there is clustering within the dataset, i.e. the extent to which observations within a child are more similar than between children, and the extent to which children's engagement within classes is more similar than between classes. The amount of variance explained by the model can be calculated using a global r^2 , where r is the correlation between actual values and the fitted values and can be thought of as the distance between an individual's predicted score and their observed score (Roberts et al., 2010). However, this method does not necessarily take into account all of the intricacies of multilevel models, and estimates may not be reliable. More complex methods exist (see Roberts et al., 2010 for a review) but these are not readily available to the non-expert statistician.

Control Variables. Control variables (age, year group, gender) were tested individually as random intercept models. In both studies, all control variables were non-significant, and with small or no effects (Table 4) so no control variables were used in further

models. The effect of age was also tested controlling for year group, but it was not significant either.

Table 4 *Fixed Effects of Control Variables in the LIS Study*

Age	Gender (baseline: Male)	Year group (baseline: Reception)
-0.01	<0.01	-0.1
CI= -0.03 - <0.01	CI= -0.1- 0.2	CI= -0.5 - 0.2
<i>p</i> = 0.14	<i>p</i> = 0.64	<i>p</i> = 0.39

Note. The models were run with restricted maximum likelihood and Kenwood-Roger correction. Negative values are coloured red for clarity.

Missing Values. One teacher did not provide children’s age, which was imputed as the average of the minimum and maximum age children could have been on the day of the visit. This was close to the average age in the sample of the same year group. The teacher also did not complete the SDQ and CBQ for the children in their classroom. Missing values are problematic and the best approach is generally considered to be multiple imputation (van Buuren, 2010). However, there were a number of difficulties that made this approach impossible in the present research. First, missing values were at level 2 in the explanatory variables, whilst most advice for dealing with missing values is at level 1 for the predictor variable. Little attention has been paid to missing values at level 2 (van Buuren, 2010), and although some methods exist for dealing with them, there are no widely available methods in the usual statistical packages. Secondly, all child-level variables were affected in the same way, for the same children, making methods for imputation less robust. Finally, other than gender there were no variables that could be used to predict the missing data.

Therefore, listwise deletion was used instead. Because of the bias this can introduce and the loss of power resulting from a smaller dataset (van Buuren, 2010), the results for the child characteristics are presented separately and as exploratory. I also took some steps to ensure the results were not overly biased: the random slope model with the complete dataset was compared to the same model with missing LIS values, and the r^2 and AIC compared. There were some differences in the coefficient values, standard errors and *p*-values of the context variables, but these differences were small (e.g. the effect of free choice decreased from 0.74 to 0.69) and so it is unlikely the deletion of data led to important biases. There

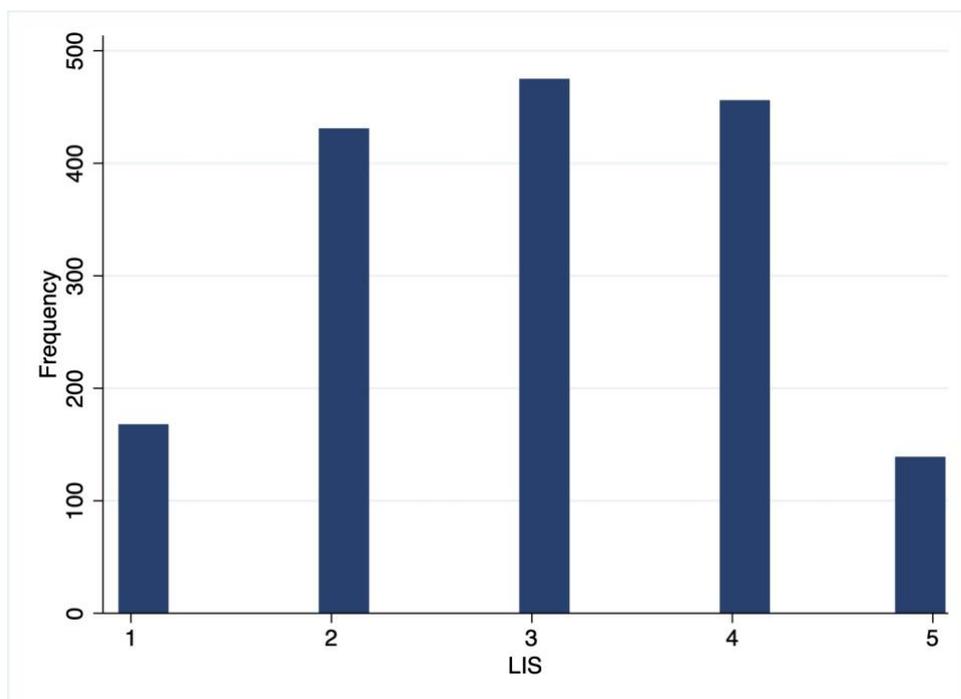
were also some differences in the AIC and r^2 : the model with missing values had a lower r^2 (hence a poorer fit), but also a lower AIC; this is likely because, with fewer children, it had fewer components and therefore it was a simpler model.

8.3.2 Results on Variability and Contextual Factors: Descriptive Statistics

The mean engagement of children was 2.98 (s.d. = 1.13), with an average engagement of 3.02 (s.d.= 1.07) in Reception and 2.91 (s.d.= 1.20) in Year 1. This is lower but within the same range as engagement in other studies using the LIS (Ulich and Mayr, 1998).

Distribution of engagement approached normality (skewness: -0.03; kurtosis: 2.2). In other words, there were a high number of ratings in the middle of the scale (2-4) and few at the extreme ends of the scale (Figure 11).

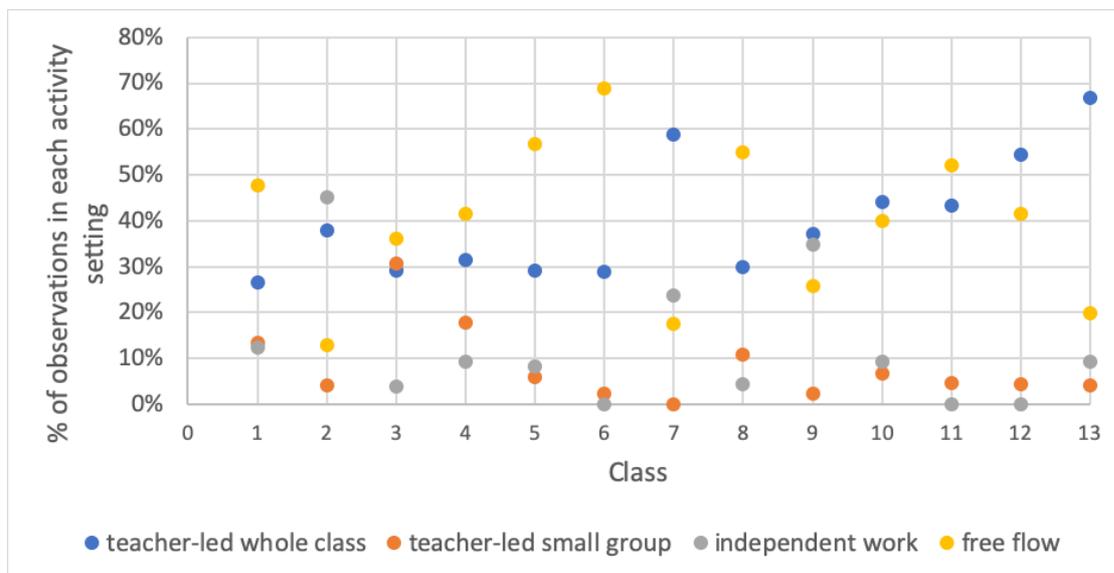
Figure 11 Frequency Distribution of LIS



Children were most often observed in whole class situations and free flow (40% and 39% of observations respectively). The children were less frequently observed in teacher-led small-group work or in independent work (8% and 12% of observations respectively). However, there were important differences between classrooms in how many observations were made in each setting (Figure 12). It should also be noted that these data represent the proportion of observations in a given setting and not the actual proportion of the day children

spent in these activities; nevertheless, because rounds were short and fairly regular, when enough sweeps were made in the class over the entire day, the number of observations is likely to be representative of the overall pattern of activities in the class. Time in whole class settings may have been slightly overrepresented as rounds were usually in shorter sequence than in other settings, but the large variation between classrooms (29%-67%) suggests that this alone cannot explain the high proportion of time spent in a whole class setting in some schools.

Figure 12 *Proportion of Observations in Each Activity Setting in Each Class*



8.3.3 Variation Explained by the Child and the Classroom Level

The ICC is the ratio of between-cluster variance to total variance and gives an estimate of clustering in the dataset at the different levels. Across all the models, only 5-7% of the variation in engagement was uniquely attributable to the child, with ICCs for the child level (within classrooms within schools) between 0.65 and 0.78 (Table 5). Confidence intervals suggested a higher margin of 13%, which is still low. In the empty model, the classroom level accounted for less than 1% of variation and the school level accounted for an additional 1.3%. However, this class effect disappeared in further models (see below) and the school effect remained low. This is likely because any variation between classrooms was already explained by the observation-level variables; this suggests that the difference in engagement between classrooms is associated with differences in the context of the activity.

Table 5 *Intraclass Correlation Coefficients for the Multilevel Models*

Model	School	School: Class	School: Class: Child
empty	0.007 CI= 0.000 - 0.370	0.020 CI= 0.004 - 0.089	0.078 CI= 0.047 - 0.128
Random intercept	0.010 CI= 0.001 - 0.077	0.010 CI= 0.001 - 0.077	0.077 CI= 0.047 - 0.124
Random slope (class level) with correction*	0.017 CI= 0.003 - 0.096	0.017 CI= 0.003 - 0.096	0.065 CI= 0.032 - 0.128

Note. Only the final model (marked *) used restricted maximum likelihood and the Kenwood-Roger correction, but differences with the model without the correction were negligible.

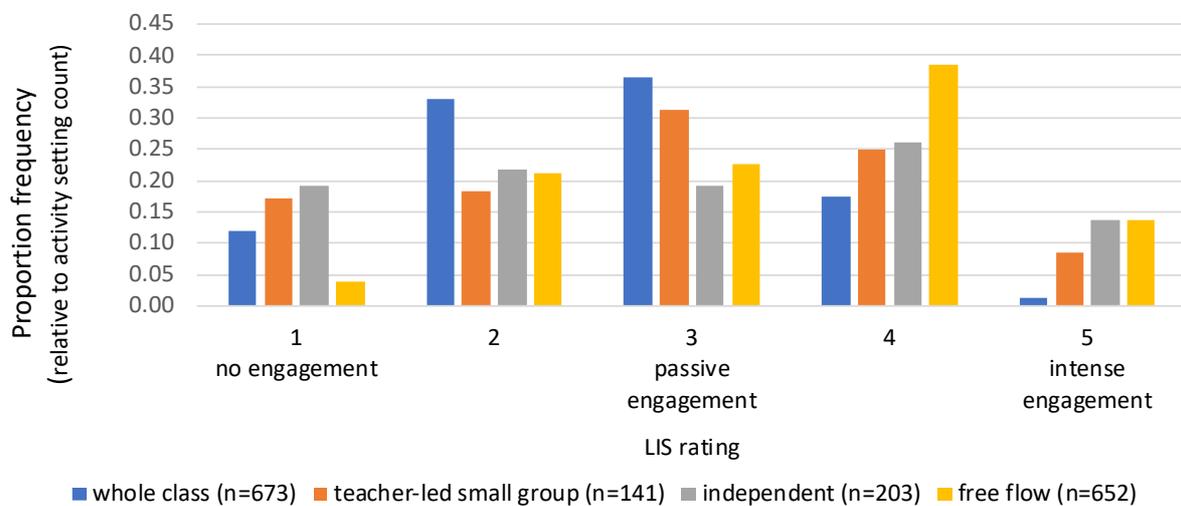
8.3.4 Variation Explained by the Activity Setting

Activity settings significantly predicted engagement (Table 6). The results refer to the random slope models which allowed predictor variables to vary at the child level. This was the best model compared to the random intercept and other random slope models allowing variables to change at the class level, based on likelihood ratio tests between the models, as well as confidence intervals of coefficients. This is likely because of the small sample size at the classroom level. Details of other models are presented in Appendix R for comparison purposes as well as in the exploratory section.

There was a significant 0.7-point increase on the rating scale when children were in free flow versus in whole class situations ($p < 0.001$). There was also a significant 0.3-point increase on the rating scale when children were in small group teacher-led activities ($p = 0.013$). However, the confidence interval for this was large, likely caused by the smaller number of observations of small group teacher-led activities (8% of observations), which makes this finding less robust. The smaller effect of independent activities (0.2 increase on the scale compared to whole class) was not significant ($p = 0.087$), though it was significant in the random intercept model ($p = 0.005$; see Appendix R). This is consistent with the pilot where only random intercepts were calculated, though the effect sizes were slightly higher in the pilot. This suggests that the difference in engagement in independent activities versus whole class found in the intercept models in fact varies with the child. Indeed, there was a small amount of variation between children in how much independent work predicted their engagement: for some children independent activities were more conducive to engagement than whole class,

but not so for other children. This was also the case for small group teacher-led activities. However, for both teacher-led small group and independent work settings, the confidence intervals were quite large. Engagement varied less by child in free flow and hardly at all in whole group settings, and confidence intervals were narrower, suggesting the effect is likely to be smaller. In other words, the effect of whole class activities on engagement is likely to be the same for different children, and the effect of free flow on engagement is likely to only vary a little between different children.

Figure 13 *Distribution of Engagement in Different Activity Settings Across All Classrooms*



Note. Distribution is represented using the proportion of ratings within each activity setting to adjust for the relative occurrence of the activity setting.

Overall, there was a large amount of variance left unexplained, as seen by the high residuals (Table 6). We can also see there was a large amount of variation in engagement within each activity setting (Figure 13). The model seemed to explain only a small amount of variance, with global r^2 of 8.5%. However, as previously noted the use of r^2 in multilevel modelling is contested (Roberts et al., 2010). Overall, the model did not violate any of the assumptions of multilevel models at level 1, though the small sample size at classroom and school levels caused some issues which makes the models less reliable at these higher levels (see Appendix S for further details of assumption checks).

Table 6 Predictors of Engagement: Fixed and Random Effects at the School, Class and Child Level

Residuals	Model parameters	Constant	Free flow	Independent	Small group teacher-led
1.0 CI= 0.9 - 1.1	Fixed effects	2.6 CI= 2.5 - 2.8	0.7** CI= 0.6 - 0.9	0.2 CI= 0.0 - 0.4	0.3* CI= 0.1- 0.6
	School-level random effects	<0.1	-	-	-
	Class-level random effects	<0.001	-	-	-
	Child-level random effects	0.1 CI= 0.0 - 0.1	0.1 CI= 0.1 - 0.3	0.2 CI= 0.1 - 0.6	0.2 CI= 0.0 - 0.7

Note. The multilevel models were run with activity setting as dummy predictor variables, and “whole class teacher-led work” as the baseline.

Fixed effects only: * $p < 0.05$, ** $p < 0.01$. Significance levels are not calculated for random effects.

8.3.6 Exploratory Results

Class-Level Random Slope Model.

Analytical Approach. As part of the sequential procedure for multilevel modelling, random slope models were also run that allowed predictor variables to vary at the class level. As explained above, these models were not as good as the child-level random slope models, and this was likely caused by the small number of classrooms ($n=13$).

Results. The child level and class level models gave a similar overall picture to the models presented above: free flow was positively associated with engagement, whereas other contextual factors were not, or not as strongly, associated with engagement (Table 7). Nonetheless, we present here a few noteworthy results, as they may highlight some avenues for further investigation, and may help to understand differences in children’s engagement.

Table 7 Fixed Effects and Random Effects for the Class-Level Random Slope Models

Model	Residuals	Model parameters	Constant	Free flow	Independent	Small group teacher-led
random slope class level	1.0 CI= 1.0 – 1.1	Fixed effects	2.6** CI= 2.5 – 2.8	0.7** CI= 0.6 – 0.9	0.2 CI= -0.2– 0.5	0.3 CI= -0.1 – 0.7
		School-level random effects	<0.01 CI= 0.0 – 0.2	-	-	-
		Class-level random effects	<0.01 CI <0.01	<0.1 CI= 0.0 – 0.3	0.1 CI= 0.0– 0.7	0.2 CI= 0.0 – 0.8
		Child-level random effects	0.1 CI= 0.0 – 0.3	-	-	-
		Fixed effects	2.6 CI= 2.5 - 2.8	0.7** CI= 0.6 – 0.9	0.2 CI= -0.2– 0.5	0.3 CI= -0.1 – 0.7
random slope class & child levels	1.0 CI= N/A	School-level random effects	<0.1 CI= N/A	-	-	-
		Class-level random effects	<0.01 CI= N/A	<0.01 CI= N/A	0.1 CI=N/A	0.2 CI=N/A
		Child-level random effects	0.1 CI=N/A	0.1 CI=N/A	0.1 CI=N/A	0.1 CI=N/A
		Fixed effects	2.6 CI= 2.5 - 2.8	0.7** CI= 0.6 – 0.9	0.2 CI= -0.2– 0.5	0.3 CI= -0.1 – 0.7

Note. * $p < 0.05$, ** $p < 0.01$

The models were run with restricted maximum likelihood and Kenwood-Roger correction. Baseline: whole class.

There was slight variation in how independent and teacher-led small group work predicted engagement depending on the class (Table 7) – in some classrooms independent work predicted engagement better than in others, and the same was true for teacher-led small group work, whereas whole class and free flow tended to be little affected by the classroom in which observations were made. However, these effects were small (0.1 and 0.2 change in the rating scale). When this variation was taken into account, teacher-led small group work no longer predicted engagement on average. This suggests that there likely were differences between classrooms in the way in which these teachers worked with children or in the nature

of the tasks, which was then reflected in the activity setting, rather than it being independent and teacher-led small group work *per se* that was associated with engagement. As before, the models did not seem to violate assumptions of multilevel models.

Association Between Engagement and Child Characteristics.

Analytical Approach. Because of missing data and difficulties in imputing values in multilevel models where the missing values are at level 2, the results presented here excluded one classroom and are presented as tentative.

The SDQ is made up of five subscales (emotional difficulties, conduct problems, hyperactivity, peer problems, and prosociality), though an overall difficulties score can be calculated by averaging the first four together. The CBQ is made up of three subscales: effortful control, surgency and negative affect.

Using the ‘bottom-up’ approach described above (Hox et al., 2010), I ran analyses first with the global difficulties SDQ score (which was not significant) then with the individual subscales. Models were run with each child characteristic variable individually and with activity setting variables. Only significant level 2 variables were added to the next step, and any variable that became non-significant was dropped from the model presented here. I initially retained variables where the *p*-value was close to the 0.05 threshold but dropped them if they were no longer close to the significance threshold in further models. In the final model, only emotional difficulties and effortful control were retained. However, it should be noted that a number of variables were correlated. For example, negative affect and emotional difficulties were strongly correlated (0.67). This suggests that in the models with a large number of variables, some of the effects of some child characteristics would have been confounded. The final model can be interpreted as the one with the child characteristics that best predicted engagement, even when accounting for all others child characteristics.

As with the multilevel modelling procedure previously described, we first fitted a random intercept model then a random slope model at child level. Initial analyses used the restricted maximum likelihood method with the Kenwood-Roger correction, but random intercept and random slope models were compared with and without the correction, using r^2 for the former and the LR-test for the latter. There was no difference in r^2 but the LR-test showed the random slope model was a better fit. Random slopes at child level of the child characteristics (i.e. models where the effects of emotional difficulties and effortful control on

engagement vary depending on the child) are not presented here as the effects were negligible and the model explained no more variation than the model presented here.

Results. Engagement was significantly predicted by emotional difficulty and effortful control (Table 8). Emotional difficulty was associated with a small decrease in engagement: 0.07 point decrease on the LIS for every 1-point increase on the SDQ 11-point scale, which corresponds to a 0.7 point decrease in engagement for a child with severe emotional difficulties (scoring 10) compared to a child with no difficulties at all (scoring 0). Effortful control was associated with a small increase in engagement: 0.2 point increase on the LIS for every 1-point increase on the CBQ 7-point scale, which corresponds to a 1.2 point increase on the rating scale for a child with the highest score (7) compared to a child with the lowest score (1). In other words, children who scored higher on emotional difficulty tended to be slightly less engaged, and children who scored higher on effortful control tended to be slightly more engaged.

Table 8 Fixed Effects and Random Effects for the Model Including Child Characteristics

Model parameters	Constant	Free flow	Independent	Small group teacher-led	Emotional Difficulty	Effortful Control
Fixed effects	2.6 CI= 2.5 - 2.8	0.7** CI= 0.6 - 0.9	0.2 CI= 0.0 - 0.4	0.3* CI= 0.0- 0.5	-0.1 CI= -0.1 - -0.0	0.2 CI= 0.1 - 0.3
School-level						
random effects	<0.001	-	-	-	-	-
Class-level						
random effects	<0.001	-	-	-	-	-
Child-level						
random effects	<0.1 CI= 0.0 - 0.1	0.1 CI= 0.1 - 0.2	0.3 CI= 0.1 - 0.7	0.2 CI= 0.0 - 0.8	-	-

Note. * $p < 0.05$, ** $p < 0.01$. Child characteristics are modelled as fixed effects, and activity setting as fixed effects and random effects at child level. The models were run with restricted maximum likelihood and Kenwood-Roger correction. Baseline: whole class.

In this model, the variation explained by the child level diminished so that about 1% of variation was explained by all three levels (ICC=0.012). This suggests a large amount of

the variation in engagement between individual children was accounted for by the variables of emotional difficulty and effortful control. The effects of activity settings were unchanged compared to the results presented earlier.

There were no important issues in the assumptions of the multilevel models, with visual diagnostics showing normally distributed and homoscedastic residuals at different levels. However, one issue that may have caused some model misspecifications is that the emotional difficulty scale (as all SDQ scales) was not normally distributed, with most values at low scores, and a long tail up to the maximum score. The relatively low sample size at level 2 variables ($n=65$) means that the Central Limit Theorem might not apply, and transformations did not change this distribution. The models were subsequently run with robust standard errors, but under maximum likelihood models (rather than restricted maximum likelihood), as the two types of models are not compatible; there was little difference in the coefficient, standard error and p -value of the relationship between child characteristics and engagement between the models with and without robust standard errors (with maximum likelihood), and so I concluded that the possible violation of the assumption was not consequential. There was a small difference in standard errors compared to the REML model, and this impacted on p -values, but this did not affect the general conclusions. However, sample size at level 2 was on the low side ($n=65$, compared to a recommended minimum of 100) for robust standard errors to work well (Hox et al., 2010).

Chapter 9 - Part II Discussion

9.1 Suitability of the LIS for the Stepping Stones Project

This part of the thesis stemmed from the potential of the LIS to be used in evaluation research as an indicator of inner motivation. However, the lack of publicly available information made further research into the psychometric properties of the LIS necessary.

The first question was whether LIS was a reliable instrument. This included establishing inter-rater reliability, which was good in the pilot and the video coding, but poor in our main's study live coding. This highlights the importance of good training and achieving a shared understanding of what the LIS is capturing. If this is not done well, as was the case when live coding, inter-rater reliability suffers. In other words, whilst it is relatively simple to be trained in using the LIS and it requires little theoretical background knowledge, it does require researchers to view children's activities through the same lens. Taking up Krippendorff's definition of reliability as the extent to which different individuals arrive at the same interpretations (Krippendorff, 2011), the LIS can therefore be reliable *insofar* as the coders interpret the classroom situation in the same way. The fact that intra-rater reliability was not total (though nonetheless high) also shows that there is some error in the way in which similar interpretations of the situation lead to different ratings. We noticed during training, for example, that once a similar understanding of the rating process had been achieved (which we did with video coding, but not live coding), disagreements were not so much in the interpretation of the event, but on borderline situations where it was ambiguous whether a child was in one category or another. Rating engagement on a 5-point scale artificially puts into categories what I defined earlier as a gradient, and so it is not surprising that these borderline cases would arise. These likely were the cause of a large number of disagreements in video coding, where 96% of inter-rater disagreements and 100% of intra-rater disagreements were within one category of each other. It is also noteworthy that despite the LIS being more subjective than on-task/off-task measures, our reliability with video coding was only slightly lower than studies using on-task/off-task measures (e.g. Kappa of 0.79 to 0.84 in Godwin et al., 2016).

An associated question was whether we could reliably capture differences in engagement across time – that is to say that any differences in engagement from one time point to another would be due to differences in our programme. I found that engagement varied hugely throughout the day, and that for any given child there was no stability from one

observation to the next. The level of variation explained by the child level was indeed relatively small, even where there were up to 30 observations for each child. However, this variability could create problems for using the LIS in interventions on two counts. First, if the number of observations for each child or classroom is insufficient, then it is not possible to build a reliable picture of this variation – in other words, assuming engagement is normally distributed around a mean (which our research suggests it is), there is a sampling issue in that too few observations will create an unrepresentative sample. Secondly, since observations happen over the course of a single day for each class, we hypothesised that events during a particular day could affect engagement; for example, teachers sometimes cite windy weather as causing restlessness. If the day of observation has an effect, it would be difficult to use the LIS in evaluations as we would not be able to dissociate the effect of interventions from the effect of the day. However, I found that overall the day had no effect on engagement: variability around average engagement was stable across days when taking all observations for each classroom. This is also in line with research in older students (Martin et al., 2015).

However, we also found that little variation was explained by class and school levels, which suggests it might be difficult to detect differences between classrooms. Nonetheless, this lack of inter-class variation may have been due to the fact that the effect was considered as an average – it may be that certain individual classrooms significantly differed. In addition, the little variation between classes disappeared entirely when the activity settings were included in the models. This suggests that it is indeed the classroom practices such as activity setting, and not structural aspects of the class and the teacher's approach, that affect engagement. This is in line with studies that have shown that process quality – that is to say the quality of interactions and other in-the-moment aspects of learning – is an important predictor of a range of important outcomes in the Early Years (e.g. Sylva et al, 2020). It also suggests that classrooms tend to vary in their use of different practices, but also that a range of practices can be found within the same classroom. This takes us to the final question I investigated.

The third question was whether deep engagement (as a proxy for inner motivation) was something our programme would be likely to influence. The large number of ratings towards the low end of the scale suggests there is definitely room for improvement, and the high variability suggested that contextual factors were important and therefore that engagement is highly malleable. Finally, the finding that a free-flow activity setting is associated with greater engagement compared to whole class is also in line with our programme aims. The finding that there was little variability between children compared to

between observations also supports the idea that engagement is easy to impact by changing in-the-moment aspects of children's learning experience, but also has implications for how engagement has been traditionally studied. I turn to this issue in the next section.

9.2 Engagement as a Characteristic of the Child or the Moment

As previously discussed, the differences in engagement between children were small, both in the pilot and in the main study. This finding is surprising in that engagement is often studied as an attribute of the child; in other words, as something that is caused by child characteristics. Some researchers even consider within-child variation to be a source of error rather than informative (Carbonneau et al., 2020). Nonetheless, there is increasing interest in intra-individual variability, which appears to have long been overlooked in engagement research, possibly because of the reliance on self-reports and aggregation of child-level data. In particular, a number of recent time-series studies by Malmberg and colleagues have shown that this intra-individual variation is in fact hugely important to understanding students' engagement across the school day (e.g. Heemskerk & Malmberg, 2020; Malmberg & Martin, 2019; Martin et al., 2015; Vasalampi et al., 2020). Even then, however, there are differences in how much or how little is explained at the level of individual children: in one study, between-children variation represented nearly 70% of variation, but in another study, most variation occurred within individual lessons. These studies, however, rely on self-reports in older children, and the study with large between-children variation asked students to report on general behaviours rather than in-the-moment engagement. It may also be that the children's age is responsible for greater individual stability in engagement; for example, Heemskerk and Malmberg (2020) argue that as children get older, their use of metacognitive strategies increases and their dependence on context may be less important. It may also be that the nature of self-reports is responsible for the difference. For example, students may report interest without displaying it outwardly, and this would allow more stable dispositions to appear. By contrast, Vitiello and colleagues (2012), using an observational instrument in preschools, found similar results to the present study, with only 7% of variation occurring between children, which supports the present findings. This research therefore contributes further evidence that attention needs to be paid to intra-individual variation in engagement, as some ways of conceptualising and measuring engagement may mask high variability between moments.

It may also have been that by using the LIS we were better able to discriminate between types of engagement which may be more sensitive to context than the distinction between on-task and off-task behaviour, which may be more closely related to children's ability to concentrate and comply with external demands. Indeed, what differences existed between children were accounted for in part by differences in behavioural characteristics of the children, namely emotional difficulties and effortful control (that is to say the ability to exert effort to control behaviour). This fits well with the well-documented relationship between engagement and self-regulation (Boekarts, 2016; Portilla et al., 2014; Stefansson et al., 2018), which comprise both emotional and effort dimensions. Compared to research with older children focusing on on-task/off-task behaviour, the children in our sample almost all had opportunities to play, where their motivation and interest may have helped them stay focused (Sansone & Thoman, 2006). This means that even children who may have otherwise struggled to comply with adult requests or with imposed academic tasks could find activities to engage them, hereby reducing differences between children in their ability to engage in tasks. This also relates to the point made in Chapter 5 that some teachers in our professional development programme (Part I) noted that some children benefitted from continuous provision because there was always something they could do. Further research would be needed to investigate this proposal, but if correct, it would suggest children with behavioural difficulties could benefit from greater freedom to choose activities for themselves and engage in play in the classroom.

9.3 Activity Setting and Other Contextual Factors Associated with Engagement

To better understand what these contextual factors affecting engagement might be, I investigated the effect of a possible candidate, activity setting. As predicted, I found that free flow was the most strongly associated with higher engagement of all activity settings. This finding is tempered by the fact that children may have differed slightly in how much they benefitted from different settings. Whole class situations were on average the least conducive to engagement, and this affected children equally. This is in line with studies with children of similar ages in the USA (Powell et al., 2008; Vitiello et al., 2012) and in older students (Downer, et al., 2007; Godwin et al., 2016; Heemskerk & Malmberg, 2020). Children are largely passive in whole group activities, and therefore it is expected that they would be less deeply engaged. However, it should be noted that children in free flow situations were usually in play situations: out of 652 observations, only 44 (7%) related to academic tasks

(reading, writing, number work), and a further 105 (16%) related to an art activity (e.g. painting, music), whilst the rest, 503 (77%), were play. By contrast, most other observations were academic-related tasks, except for a few occasional whole class games. Therefore, it may be that it was play rather than self-directedness which fostered greater engagement in free flow settings. In this case, it would suggest that playful activities are beneficial to children's engagement; even so, this type of practice is also aligned with the programme's objective.

In addition, children were slightly more likely to be engaged in small group settings with a teacher than in whole class settings, but this varied depending on the child, and possibly on the classroom. The presence of a teacher means the child could be externally regulated: any distraction or sign of being 'off task' is quickly addressed and the adult can refocus children on the task at hand. This would reduce opportunities for events to be scored on the lower end of the scale. However, this alone cannot explain the high incidence of observations at the high end of the scale. It may be that in smaller group situations teachers are better able to individualise and stimulate learning to make it more personally meaningful to children, or that the teacher's attention is in itself stimulating as children want to do their best. This would also concur with the exploratory analyses at the class level, where we found that the effect of teacher-led small group work on engagement was no longer significant when it was allowed to vary with the class, and its effect varied much less by child. This would suggest that children's engagement in small group teacher-led work depends on classroom-level characteristics, for example the teacher's skills and interpersonal style, and the sorts of tasks the teacher usually sets. It may also be that there was more variability in what teacher-led small group work looked like in different classrooms, whereas free flow or whole class work often had similar characteristics across different classrooms. For example, small-group work can be an opportunity for interactions that deepen and extend children's thinking (Siraj-Blatchford, 2010), but research by another member of our team has shown teachers tend to focus small-group work on basic skills instruction instead (Eberhart, 2020).

Children in independent work were also slightly more highly engaged, on average, than in whole class settings, but this association was not significant, perhaps because the nature of tasks was more important than the fact that work was carried out independently. The effect of independent work is less clear in other research as well. For example, Heemskerk and Malmberg (2020) found that individual work was positively associated with engagement compared to whole class instruction, but Godwin et al. (2016) did not find an effect. It has been proposed that older children may be better able to stay focused during

individual work (Heemskerk & Malmberg, 2020). It should also be noted that other studies tend to consider peer work and individual work separately even though instances of peer work tend to be rare (e.g. Heemskerk & Malmberg, 2020), whereas in this study the two categories were aggregated.

In summary, overall this research strongly suggests that helping teachers focus on children's agency in learning will improve engagement, for example by increasing opportunities to direct their learning (as in free flow settings) or providing support and stimulation (as in small group work). However, there are also a number of other classroom processes which foster or inhibit engagement, as suggested by the unexplained residual variance, and these merit further investigation. It is these possible mechanisms for engagement I discuss next.

9.4 Mechanisms for Engagement

If activity setting only explains some of the variation in engagement, and if children's personalities and the classrooms they are in explain so little, then the question remains about what it is exactly that affects children's engagement. What is it about classroom activities that engages children at any given point in time? Clearly, self-directedness plays a role, but the large amount of unexplained variation indicates there is more to it. This is not particularly surprising, considering the wealth of research studying different antecedents and facets to engagement (as reviewed by Boekaerts, 2016), but is an important reminder of the importance of each classroom moment. In particular, research on teacher-child interactions suggests that this is likely to be an important factor affecting engagement. For example, Downer et al. (2007) have shown that students at risk of school problems are more engaged in whole class instruction when in classrooms with a positive climate. Not only this, but classroom quality was a stronger predictor of engagement than the focus of instruction (basic skills or inference) (Downer et al., 2007). The importance of teacher-child interactions and classroom climate is also supported by research on teacher sensitivity and autonomy supportive styles (e.g. Skinner & Belmont, 1993; Reeve et al., 2004). Research also supports a bidirectional effect between children's behaviour in classrooms and teachers' interactions with them: teachers tend to be less warm and sensitive towards children who disrupt lessons or are more often in conflict with teachers, which could lead to the child's further disengagement (Skinner & Belmont, 1993; Portilla et al., 2014).

Children's interactions with peers also likely play a role, as they can act both as distractors and as facilitators of learning. The role of peers in engagement is unclear; whilst some studies with older Primary school students have found small-group work to be more conducive to engagement than individual work (e.g. Downer et al., 2007), this finding is not shared by others studying the same age group, albeit in an English rather than US context (e.g. Heemskerk & Malmberg, 2020). One study in preschools found that the association between engagement and the presence of peers was different depending on whether the activity was academic (positive association) or play (negative association) (Powell et al., 2008). The impact of peers on engagement is therefore likely to itself be under the influence of multiple other factors, such as the nature of the children's relationships, the nature of the task, the purpose of peer engagement, how the teacher protects children from distractions, etc. In this research, we did not look specifically at the impact that the presence of peers had on engagement, but noticed that, as the research suggests, they could be both a source of inspiration and motivation (for example when children investigated marble runs together after one boy initiated the game) as well as a distraction.

Finally, the content of tasks is also likely to be responsible for differences in engagement. Previous research has shown that students are less likely to be engaged in tasks involving basic skills instruction rather than instruction focused on higher order thinking skills (Downer et al., 2007). Children's interest in the tasks is also likely to play a role (Patall et al., 2016; Renninger & Bachrach, 2015). For example, characteristics of tasks such as the challenge they present, or whether they are hands-on or novel might trigger children's interest (Renninger et al., 2018; Fredricks et al., 2004). From this perspective, as previously discussed, self-directedness is not only conducive to engagement because of the sense of autonomy and choice it provides, but also because it creates more space for interests to emerge and be sustained. It would also account for the classroom effects of teacher-led small group activities and independent activities – those teachers and activities that are able to trigger and maintain interest will generate sustained engagement. For example, in one classroom the teacher was often using open-ended or creative tasks for individual work. In one lesson, the children had to make small worlds using a variety of 3D shapes, with the teacher emphasising the vocabulary and properties of the shapes in conversations with them. In another lesson, the children were making repeating patterns using beads and blocks and whatever materials they wanted. Possibly as a result of this, children tended to be much more engaged in independent tasks in that classroom compared to other classrooms, with an average of 3.7 (based on 45 observations), compared to the average of 2.9 for independent

work for the entire dataset – this is well above the 0.2 random effect of independent work, and such an outlier may partly explain the large confidence interval. This also shows the potential of well-designed independent tasks for children’s engagement: in that particular classroom, engagement in independent work was even higher than in free flow.

It would be interesting to see whether the LIS ratings overlap with indicators of interest – that is to say whether the LIS does partly capture interest – and whether known triggers of interest are a good contextual predictor of engagement. Indeed, conducting this research has raised a number of questions around what the LIS is in fact capturing, and next I examine this and related issues in the measurement of engagement.

9.5. Further Issues in the Measurement of Engagement

Because the aim of this research was, at heart, to critically examine the LIS, it is important to not only discuss its reliability but to frame it within a broader discussion around the validity of such tools and how they are used. Conducting the present research raised questions around the meaning of the ratings and results such studies produce, which I now critically examine.

9.5.1 Validity Questions in the Face of Measurement Error

One question which arose from this research pertains to differences in the instrument’s validity depending on the situation. For example, engagement in whole class situations was more difficult to infer than in free flow, because of the child’s passivity. This may have increased errors in whole class settings, for example by creating an artificial ceiling to our ratings; in other words, children may have been more likely to be rated as ‘passively engaged’ because they had little opportunity to express the depth of their engagement rather than because they were only superficially engaged. Another way to think about it is when we attend an interesting conference or lecture, where our cognitive engagement may be high, even our enjoyment, but with little outward expression. Whilst every measurement will have some measurement error, and particularly so when we are aiming to capture internal states, statistical methods assume that measurement error is random and not correlated with any of the variables of interest. Therefore, if whole class situations are associated with greater measurement error at higher ends of the scale, this would put into question the finding that whole class situations were less conducive to engagement than other settings, as well as the validity of the LIS. However, there are a number of factors which suggest this problem is

likely not a serious threat to the validity of this research. First, in many whole class situations, it was not so much a case that there were few opportunities to express engagement, as much as there were few opportunities to engage at anything but a routine level. This included listening to instructions, chanting letters and sounds, copying words from the board, waiting for another child to give an answer: these were frequent occurrences where there was simply no opportunity for the child to exercise higher-order mental capacities, and therefore the lower ratings were in line with the descriptions of the LIS, rather than it being an artefact of the observation process. Secondly, there were occasions where we did observe high engagement in whole class activities, which shows that a lack of opportunity to express deep engagement was not a characteristic of all whole class situations. Finally, the children's young age, whilst it creates challenges for verbal modes of inquiry, creates affordances when it comes to observation tools. Their relative lack of self-control means they will find it more difficult to control restlessness from boredom, and it will be easier to observe. Young children are also more effusive in their enthusiasm, which again makes it easier to observe, and their use of self-talk also makes it easier to understand what might be going on for them.

However, individual differences come into play here. If some children are more likely to display outward signs of their inner states than others, then it means the LIS will differ in how accurate ratings are depending on the child, which is problematic. Not only this, but similar levels of engagement might be expressed differently depending on the child, and therefore individual behaviours will have different meanings depending on the child. For example, the excitement of an otherwise quiet child might have more weight than an excitable one, but it might be more difficult to gauge the depth of their engagement at other times. It is for this reason that some researchers have argued that the LIS is best used by people who know the children well. In the present research, we did not know the children prior to visiting the classrooms. Nevertheless, I argue it is possible to infer children's internal states reasonably well. The large intra-individual variability is testament to the fact that individual children displayed the whole range of the engagement gradient – it was not the case, for example, that some children were always rated as passive because of their way of responding to events. In addition, after spending even a small number of days in a classroom observing children in such detail, as was the case in this research, it was easy enough to form some impression of their personalities. It may be that in the future, a period of familiarisation is necessary, so that the researcher has a better understanding of these individual idiosyncrasies in how different children express their engagement before rating begins. In addition, it may be that teachers are able to contribute to data collection by using the LIS

themselves in classrooms. They may be better placed to interpret and put children's behaviour into context, and therefore may be able to make more accurate as well as faster judgements.

9.5.2 Face Validity – What Are We Capturing?

A further question which arose out of the research, though it was not directly addressed in the research questions, concerns whether the LIS has face validity for the purposes of the Stepping Stones project – that is to say, whether the instrument adequately captures inner motivation. Although intrinsic motivation is part of the description of the LIS, it is impossible to observe it directly and it must be inferred from the situation. As with engagement, different situations vary in how easy it is to make this inference. In particular, passive engagement and compliant (but bored) engagement may outwardly look very similar (Fisher, 2011; Zyngier, 2008).

Secondly, even if we can adequately infer internal states, this might tell us little about the reasons for engaging in the activity. That is to say, one may be deeply engaged at a cognitive level, but only for extrinsic reasons, and therefore not interested or intrinsically motivated. Sometimes it is possible to make these inferences. This is the case in some of the training videos, where a child is doing mathematical computations within a limited time frame. The child's apparent nervousness about doing well with respect to other children suggests that his motivation is at least partially a controlled form of motivation, despite the child being very committed to his task. But in many cases, it can be very difficult to tell – and in the case of the child above, he still received a high rating because of his dedication and focus. This means a child can receive a high rating despite not being internally motivated, which challenges the idea that the LIS is an indicator of inner motivation. This question is also related to the relationship between engagement and children's characteristics, such as effortful control. A child with poorer self-regulation skills may be internally motivated but not able to regulate themselves enough in order to maintain focus and achieve their goal. Vice versa, a child may not have been very interested, but expressed a high degree of commitment because of externally motivated reasons, such as wanting to do well.

This relates to the nature of the LIS as a global scale, which means different facets of the construct are necessarily aggregated. One consequence is that ratings represent different realities depending on the situation, even if we assume they can be reliably inferred. For example, whilst engagement was on average higher in free flow than in whole class, there

were nonetheless a high number of observations at the lower end of the scale. But the way this lower score was expressed was different in each setting. Children at the low end of the scale in whole class situations usually exhibited signs of boredom, including restlessness, wandering eyes, or disruptive behaviour. However, children in free flow seemed rarely 'bored' with the task at hand and their lack of engagement usually took one of three forms: either they wandered aimlessly, they flitted between tasks and never became fully engaged in any one of them, or were distracted by what other children were doing around them. In free flow, passive engagement often involved children chatting and having rich social lives whilst being 'busy' in an unrelated activity, like drawing. As a result, the same ratings in different settings sometimes *felt* very different to us as raters - indeed, together we discussed how some events made us feel uncomfortable. Clearly, therefore, something was happening which the LIS was not capturing. This implies that the instrument is not able to differentiate between these important differences in children's experience and opportunities for learning. In addition, it should also be kept in mind that whilst engagement is important, some activities may be valuable for other reasons. For example, social interactions often decreased ratings because they signalled attention was split, but they would have served other purposes for the children, such as forming friendships and may have been intrinsically motivated despite children engaging only superficially. Likewise, free choice settings offer opportunities to practice planning and task switching, which may result in more fragmentation of activity and therefore lower ratings.

These differences also entail very different lessons for what teachers should do. Whilst any low rating might cause a teacher to be concerned, a child being aimless in free flow will need a very different type of action compared to a child bored on the carpet in a whole class setting. I would argue that the former provides much more opportunity for the child to engage positively and in an agentic way, and it may be the teacher only needs to provide timely support to help the child re-engage. Therefore, the presence of an adult during free flow might help increase engagement. For example, an adult might be able to make suggestions or provide the child with needed support (e.g. Perry, 2013), which we observed on occasion. The adult would also be able to join in and sustain children's play, a practice which has been shown to be hugely beneficial to learning (Bodrova & Leong, 2018; Siraj-Blatchford & Sylva, 2004) but which we hardly ever observed. Boredom in a teacher-directed activity such as whole class work, however, may only be improved through external regulation (the teacher refocusing the child or extrinsic motivators), or by the teacher

changing the task to make it more interesting or appealing (a strategy which would also apply to other settings).

Finally, these questions around the validity of the LIS in capturing agentic engagement and intrinsic motivation raise broader questions about looking at children's activities, and in particular children's play, through an adult's lens. Because training others in the LIS required articulating very explicitly how to interpret different behaviours, and discussing any differences in our interpretations, it also made us reflect on the way in which our own values influenced what we noticed and how we understood children's actions. One clear example where our adult-centric view sometimes caused difficulties was in the case of weapon play. From an adult perspective, it sometimes felt chaotic, even violent (though it never led to actual violence), and we sometimes frowned upon the repetitive themes and actions - whereas repetition in the case of exploration or academic skills was not viewed negatively. Looking more closely, however, it was possible to see creativity, rich storytelling and cooperation, but this required us to push past our own instincts around how children ought to play 'nicely'. Indeed, whilst it was sometimes difficult for us to acknowledge these children as deeply engaged (rather than frivolously or chaotically engaged), research suggests weapon play and fighting themes should not be dismissed or discouraged (e.g. Broadhead, 1992; Holland, 2000).

9.6 Limitations of the Present Research (Part II)

This research has a number of limitations. Whilst some have already been mentioned throughout this chapter, here I discuss some relevant issues further.

9.6.1 Sample Size

First and foremost, the findings around predictors of engagement was limited by the small sample size at the child level and at the class level. The sample size was limited by my capacity, in the time left after the pilot and Part I of the thesis, to conduct the fieldwork, as well as the constraints of the COVID-19 pandemic. This caused issues in the random slopes, and in particular in terms of large confidence intervals for some effects, and non-normality of residual variance in independent work random effects. The latter was also caused by large deviations from the fixed effect by a small number of children and classrooms – in other words, given a small sample size, the effect of outliers was stronger and this created issues in confidence intervals. Whilst this creates issues in modelling, which is why these results were

presented as tentative, it is also an interesting finding in and of itself. In the future, given more time, it would be interesting to study what may have been different about these children and classrooms to cause such deviations (as I have done above with the classroom with high ratings in independent work), rather than treat them as statistical anomalies.

By contrast, the sample size at observation level was sufficient to produce reliable fixed effects. Similarly, the confidence intervals around inter-rater reliability estimates suggest the sample size was sufficient for this area of investigation.

9.6.2 Non-Random Sampling and Generalisation

A second issue concerns the representativeness of the sample of classrooms. I used purposive sampling because other methods were unfeasible, which means the schools were likely not representative of all schools in England. In particular, children came from largely white, middle class backgrounds, though a number of schools also had pupil deprivation premiums at or below national average. This lack of representativeness poses issues for generalising the conclusions, though the replication of the effects across two different studies (the pilot and the main study) gives greater validity to any generalisation. In addition, care was taken to create a wide-ranging sample of schools, that is to say to target schools that used a wide range of pedagogical approaches and with different characteristics. Nonetheless, important differences between individual classrooms may have had a strong influence on the results, and in particular in random slopes given the small sample size, as described above. Whilst limiting the validity of some of the results, it also opens up interesting questions and shows that whilst between-classroom variation was small, some things are happening at the level of the classroom which clearly make a difference.

Similarly, whilst children were randomly selected, for 8 out of 13 classrooms the children had to opt-in and this means that there may have been in bias in terms of which children participated. For example, children whose parents worried about their child's disruptiveness or those who are weary of scrutiny might not have wanted their child to take part.

9.6.3 Choice of Predictor Variable

An additional limitation of this study which needs to be considered is the categorisation of context into different activity settings. Whilst having advantages in terms of simplicity of coding, the different settings encompassed a wide range of activities which

likely hid more complex phenomena. For example, it may be that playtime when children have been otherwise taught formally is used for rest and release rather than focused play compared to settings where children spend long periods of time in free flow as part of their classroom routine. However, further splitting activity settings would have resulted in many categories with few or no observations in some classrooms, which would have created issues for analysing their impact. I also considered investigating subject-specific engagement, but it is likely to vary by child and there were not enough observations for each subject for each child to investigate this adequately. However, some types of activities did seem to be more conducive to engagement than others; for example, in some classrooms children seemed to be more engaged in story time than in other whole class situations (an average engagement of 2.9 in story time versus 2.6 for the general whole class average). It may be possible to create more fine-grained categorisations, but this was not feasible in the present research given the additional work in terms of reliability and validity creating such a coding scheme would have entailed. There may also have been different ways altogether of categorising context. For example, some initial analyses were run with a different coding scheme, with activities coded as having been freely chosen, as containing some amount of choice or as offering no choice at all. The results were similar to those presented here, but I decided not to pursue these analyses further because of the high collinearity with activity setting ('free flow' and 'choice' were indistinguishable) as well as the greater ambiguity in the choice categories. In addition, it may be that what matters is the way in which children are given freedom and support in a self-directed setting, rather the activity setting per se (Eberhart, 2020). In the future, this may be an important area to investigate further.

Finally, one limitation of this research, which has been the common thread throughout this chapter, is that the LIS, as all measures, is imperfect. The essence of this chapter was to investigate the extent to which this poses serious problems to research using the LIS. This is the note on which I therefore conclude this chapter.

9.7 Concluding Remarks to Part II

Despite some limitations, the present research provides strong evidence that the LIS would be a good instrument to use in the Stepping Stones project. The LIS tells us something about children's experience of learning that is richer than measures of engagement that focus on on-task and off-task behaviour, and it allows us to capture differences in both the amount

and the quality of engagement. In a project such as ours, it would provide an important source of information about the impact of classroom practice on the children, and the present study suggests that it would do so reliably enough for this information to be useful. Good inter-rater reliability can be achieved given sufficient training, and whilst engagement is highly variable, it does not seem to vary from one day to another within each classroom.

Nonetheless, the LIS has a number of weaknesses and so these need to be taken into account when using the instrument. It requires inferences to be made, which makes inter-rater reliability a delicate endeavour; however, that is also its strength because it allows us to use our full human capacities and our understanding of the situation to give a rating. Because it is a global scale, it provides only a coarse snapshot of children's experience, and so interpretation of ratings needs to be done with care. In particular, the present research raises some questions around the relationship between involvement and inner motivation. Given the prominence given to concentration as a signal of involvement, and the possibility of children being highly focused for external or introjected reasons, high ratings of the LIS do not necessarily equate high inner motivation and there is a risk that simplistic interpretations of the ratings would lead to wrong conclusions about children's motivation. Similarly, children may be engaged in intrinsically motivated activities but in superficial ways, such as resting or socialising, which again would not be captured through the LIS. Not only does this mean that research findings need to be interpreted cautiously in terms of what differences in the LIS means for children's experiences of learning, but these limitations also raise questions about the extent to which the LIS overlaps with and therefore captures inner motivation.

To counter these weaknesses, other measures and research methods are likely to be needed to provide a fuller picture of children's experience of teaching practices and in particular their inner motivation to engage in learning.

Chapter 10 - Overall Conclusion

What if we built a better world and it didn't raise test scores? (Anonymous)

10.1 Overview of Key Findings and Contributions

Overall, this thesis makes a contribution towards our understanding of children's autonomy and inner motivation in the early Primary classroom and teaching practices that support these, as well as how we may be able to study inner motivation in classroom contexts. Specifically, it provides strong evidence that teachers are hindered in their attempts to increase children's autonomy by the larger systems around them. It also provides a new model for understanding the tensions teachers need to navigate between teacher control and children's autonomy, as well as contributing to broader discussions on autonomy, agency, motivation and engagement.

In Part I, I used interpretive methods to understand teachers' attempts to provide greater opportunities for children's autonomy through a professional development programme in the form of a Community of Practice. Through the programme, participating teachers were encouraged to take an inquiring stance to their practice and experiment with different strategies to promote children's autonomy. Through stories of change, I presented a contiguous analysis of teachers' practice before and during the programme, with a focus on the strategies they experimented with. Using thematic analysis, I presented a model describing teacher-directed practice as an ecosystem of teacher control within which teachers provide pockets of space for children's autonomy; I argued that this provides a more accurate and nuanced understanding of the tensions between teacher control and children's autonomy than models describing paradigm shifts and a balance of practices. In Chapter, 5, I also explained how teachers' struggle to increase children's autonomy was exacerbated by top-down pressures from the accountability system and the need to meet strict learning objectives and outcomes, with senior leaders mediating those pressures. I also examined tensions teachers experienced between teacher control and children's autonomy, as well as attempts to increase autonomy which teachers described positively. I concluded on the need to promote not only children's autonomy but also their agency by attending to children's need for *support* and *stimulation*, in order to foster children's inner motivation and feeling of competence. I argued that these aspects of children's learning were mutually reinforcing, rather than autonomy unidirectionally leading to inner motivation. I concluded Part I with a

discussion on how teachers' process of inquiry for increasing autonomy in the classroom may be better supported, drawing on research on teacher learning.

Part II was concerned with measuring inner motivation for research purposes and focused on the validity and reliability of an existing instrument, the Leuven Involvement Scale (LIS). This instrument aims to capture a form of engagement in learning activities that is conceptually related to inner motivation. I investigated the reliability and stability of the instrument using statistical methods, as well as factors associated with variation in engagement using multilevel modelling. I found that the LIS can be reliable as long as raters share a common understanding of different behaviours, and that a single rater can be consistent over time. However, I also discussed that in many situations the LIS did not fully capture intrinsic motivation. Instead, it ignored superficial engagement in intrinsically motivated activities and resulted in high ratings for deep but externally motivated engagement. Whilst the LIS is a useful tool, this has implications for the interpretations to be made of any findings using the LIS. In addition, I found that engagement varied hugely from one moment to the next, with very little variation between children, though the day of observation did not have an effect. What little variation existed between children was partly explained by individual differences in emotional difficulties and effortful control, which are both likely to be related to children's ability to self-regulate and remain focused on tasks. However, this research suggests that it is the individual moment that matters most, rather than characteristics of the children. To better understand the influence of contextual factors, I investigated the association of engagement and activity setting (whether children are in teacher-directed whole class, teacher-directed small group, independent or free choice situations), and found that children were significantly more engaged in free choice settings compared to whole class teacher-directed settings. However, there was a large amount of remaining variation and I discussed the implications this has for the role of teachers in supporting children's engagement. In both parts of this thesis, I also considered the limitations of the methods used.

10.2 Synergies Between Parts I and II

The two parts of this thesis were distinct in many ways: they focused on very different aspects of classrooms (teachers' change in practice versus children's engagement), they drew largely on different bodies of literature (pedagogies and teacher learning versus psychology of education) and used very different methodologies (interpretive versus statistical methods).

However, they also had a number of overlaps, in that they were both anchored in Self-Determination Theory and were ultimately concerned with children's inner motivation to learn, as well as being part of the same overall project. In addition, it is possible to see some similarities and synergies in the findings, which I outline below.

10.2.1 The Impact of Free Flow Versus Whole Class

In Part II, I found that children were significantly more engaged in free flow than in whole class teacher-directed situations. This may have been partly caused by the fact that children in free flow were often playing (rather than doing academic work). In addition, I found that there was also a large amount of variation in engagement within each activity setting. I discussed the role that teachers may be able to play in engaging children in free flow activities, as well as how the nature of activities and the extent to which they triggered children's interest would also influence children's engagement. Taking into account the observations of Part I, however, gives new meaning to these findings. Free flow settings (or what the teachers in the programme called continuous provision) provide space for children's autonomy because children have opportunities to pursue activities that interest them. However, without support or stimulation (which could come from the environment as well as peers and the adults), children may struggle to experience agency in free flow situations. In addition, peers may distract children or power dynamics may hinder children's agency, even in the presence of autonomy. By contrast, whole class situations are theorised to lead to less engagement because they either lead to boredom and disengagement, or to passive compliance. However, whole class situations need not be focused on drill. They can be opportunities for dialogue, meaning making and community-building, which could support children's feelings of autonomy. They can also be opportunities to trigger children's interest and increase motivation and engagement. Therefore the conclusions from Part I on space, support and stimulation could provide avenues both for teachers and researchers to explore the affordances of different activity settings, as the same activity setting might differ with respect to these three 'pillars' of agency.

10.2.2 The Influence of Policy

Secondly, both parts of this thesis relate to the influence of policy, both national and school-level. In Part I, this was through the ways in which accountability pressures and learning objectives as well as school-mandated practices hindered teachers' ability to increase

children's autonomy in the classroom and reinforced the ecosystem of teacher control. In the context of changes to educational policy in England, Part II suggests that an increase in the emphasis put on teacher-led and whole class practices may lead to fewer opportunities for children's deep engagement. This thesis suggests these policy influences are likely to have an impact on teachers' choices in the classroom as well as children's autonomy and engagement. It also suggests that in Part II teachers' use of free flow versus whole class or other activity settings may not have been under their own control, but mandated from above.

10.2.3 The Place of Play in Schools

Both parts of this thesis provide some insights on the place of play in schools in England and its relationship to learning opportunities. In both cases, there seemed to be an under-realised role for play in supporting children's learning, though in opposite ways. In Part I, the emphasis on learning objectives meant that children's play was often instrumentalised and limited (see also Wood, 2019), when it was allowed at all. By contrast, in Part II, which included a large number of Reception classrooms, we saw that children rarely engaged in academic activities in free flow. There may be a number of reasons for this. It could be that in the classrooms we visited there were few academic activities available in free flow for children to access independently, as might be found in a Montessori classroom (Lillard, 2017) or such as the opportunities for exploration that Claire and Julie provided in their continuous provision. It could be that teachers did not capitalise on children's play as opportunities to learn important academic concepts, for example around numbers and shapes (Williams, 2021), as we rarely observed teachers engaging with children in free flow – and this itself may have been driven by the need for teachers to focus on learning objectives through small-group work (Pyle & Danniels, 2017). Finally, it could be that children prefer non-academic activities in free flow, either because they have come to associate academic learning with teacher-directed work (Howard, 2002) or because of a lack of interest. The first two explanations for children's lack of academic engagement in free flow suggest that teachers may need to pay greater attention how academic learning can be better supported in free flow (e.g. Ephgrave, 2018; Fisher, 2016; Jensen et al., 2019; Lillard, 2013; Mardell et al., 2019; Pyle & Danniels, 2017; Williams, 2021; Wood, 2019). By contrast, the latter explanation would provide arguments for those advocating a greater focus on teacher-directed activities to ensure various learning objectives are met. Therefore it may be important to investigate this issue further.

10.2.4 The Importance of Attending to Children's Experiences

This thesis is anchored in Self-Determination Theory, which conceptualises autonomy as a feeling of being in control and engaging willingly in activities. Whilst in Part I we did not explicitly study children's experiences of the programme, the idea of spaces for children's autonomy denotes the idea that in order to experience autonomy, children need opportunities to do so. However, we also discussed, through teachers' experiences and our observations, that these spaces are not enough, as children also need to have agency to act on those opportunities. Part II, by contrast, explicitly investigated children's experience by focusing on the LIS. In other words, both parts of this thesis emphasise the importance of attending to children's experiences of learning activities and classrooms.

10.2.5 Relatedness and Community

Finally, one common limitation of both parts of this thesis is that neither attended to the social nature of learning. In Part I, by focusing on autonomy, we may have ignored the other two components of Self-Determination Theory: competence, and relatedness. Whilst we highlighted the importance of children's competence for them to be able to act on opportunities to lead their learning, we have not discussed the importance of relationships and belonging in fostering children's ownership of their learning and in their endorsement of tasks (Ryan, 2017). However, research on learning communities suggests this is a highly important aspect of children's agency in learning (Hofmann & Rainio, 2007; Seigel, 2001; Thomas, 2021). In addition, we have noted that teachers often had warm relationships with children and most teachers in the programme were sensitive in their interactions with their students. This may provide a useful 'way in' for us to connect the practices we promoted with teachers' existing practice, and support teachers' feelings of self-efficacy by highlighting their existing proficiency. It may also help us to highlight the importance of classroom climate for the success of autonomy-supportive practices, such as cultures of trust, risk-taking and dialogue such as those found, for example, in democratic settings and Reggio Emilia classrooms.

Similarly, by focusing on individual children with the LIS, we may miss how it is influenced not only by the presence of peers (e.g. Downer et al., 2007; Powell et al., 2008) but by dynamics with specific individuals (e.g. friendships) and by the broader learning community and classroom culture.

10.3 Implications of This Research for Teachers, Researchers and Policy Makers

This thesis was concerned with teachers' support of children's inner motivation and autonomy in the early Primary classroom in England. As I have shown, teachers' practice is itself under the influence of higher-level decisions, in particular mandated practices and the high accountability regime in place in England. Therefore, this thesis has implications not only for other researchers interested in similar issues, but also for teachers and policy makers. Before concluding this thesis, I therefore describe what these implications might be. What they have in common is responding to the call by educator Susan Harris Mackay to "turn down the volume of our adult agendas" (2021, p. 162).

10.3.1 Recommendations for Researchers

Sayer (1997) argues that critical social sciences are weakened if they cannot provide feasible alternatives to the models they are critiquing. Whilst the project has been focused on developing such alternatives, the difficulties the teachers faced mean that they made few changes compared to the status quo. In addition, a main lesson from the thesis has been that fostering autonomy and inner motivation is complex and emergent, and that it cannot be determined by a clear set of rules. This perhaps leaves us with little clarity as to what a feasible alternative might look like, assuming a transition to one would be possible. Here, I suggest that looking to alternative pedagogies such as democratic settings would be fruitful for researchers and educators, not as gold standards to be replicated, but as feasible alternatives from which we can learn, in particular about how tensions are resolved in practice. In addition, we have noted many of the strategies in our programme come from Early Years pedagogies; indeed, Early Years practitioners have a long history of successful child-led practice. The application of Early Years practices to older children may provide an interesting avenue of further research.

As well as refining what practices might foster autonomy and agency, there is a need to consider how teachers can be best supported in achieving this, in particular within a system that greatly restricts their ability to do so. In addition, programmes aiming to increase children's autonomy, such as the Stepping Stones project, will need to help teachers negotiate the tensions between different goals and values of education and ways of working. This will require helping teachers make their assumptions and beliefs visible and exploring uncomfortable feelings as well as encouraging deep meaning making. This requires spaces

where teachers are able to be vulnerable and question their practice, as well as giving them ownership of the inquiry process and potentially attending to new skills they will need.

Finally, this thesis also shows that viewing educational practices through the lens of children's experiences and responses is not only valuable as a research tool, but is also necessary to ensure we go beyond adult-centred goals of compliance and learning objectives, and instead create understanding to support joyful, internally motivated and agentic learning. This may also require further research into issues of power in education. It may also be that, as well as more qualitative research, we also need different measures and tools to guide such investigations. Whilst the LIS has many advantages, it also has limitations. Further tools may need to be developed, for example classroom observation instruments specifically targeting autonomy-supportive or democratic practices.

10.3.2 Recommendations for Teachers

This thesis is largely about teachers' practice, and therefore has important implications for teachers. However, questions around how to support children's autonomy in the moment cannot be adequately answered in theory; they need to be worked through by individuals as the answer will vary depending on a number of factors, including the child themselves and the conceptions held by the adults. What theory can do is provide 'beacons' for illuminating moment-by-moment decisions, that is to say providing ways of thinking about these problems. In teaching practice, such ponderings may help teachers develop praxis (theory-in-practice). In particular, I argue that we need to help teachers develop an inquiring stance towards their practice so that they can tune in to children's autonomy and agency in the moment. This will allow them to become responsive and attend to opportunities children have for *space*, *support* and *stimulation*. This inquiring stance may require teachers to ask non-judgemental, diagnostic questions of their practice, such as those proposed in Figure 14.

Figure 14 *Diagnostic Questions for Teachers to Support Agency in the Classroom*

- What do children need, and what are the obstacles, for them to be agentic in their learning?
- Are activities meaningful to children? Are children engaging willingly? Are they feeling interested, curious, motivated? Do they care/ does it matter to them?
- Do they know what to do? Do they have the right skills and tools? Do they need additional or different support?
- What opportunities are there for children to have initiative or to have influence over the learning? How *much* space do children have for making the learning their own?
- Are children's views and strengths valued – by adults and by other children? How can the classroom community better help children feel valued, heard, understood?
- Who has power in the classroom (both adults and children)? Who makes decisions and who follows? What needs to change for children (specific individuals or in general) to feel empowered?

10.3.3 Recommendations for Policy Makers

Given that our educational institutions are so deeply invested in a banking system, teachers are more rewarded when we do not teach against the grain. The choice to work against the grain, to challenge the status quo, often has negative consequences. And that is part of what makes that choice one that is not politically neutral. (hooks, 1994, p. 203)

An important contribution of this thesis is to show how the ecosystem of teacher control and pressures from outside the classroom stop teachers from experimenting in their practice and makes it impossible for them to fully engage with practices that support children's autonomy and agency in the classroom – space to follow interests, to diverge from the learning objective, to have initiative and influence, and to discuss and disagree.

Autonomy and inner motivation are likely not important considerations for current policy makers in England, and complex, nuanced conceptions of learning are unlikely to have

much persuasive power over those who are influenced by easily measurable outcomes. And yet, without a change in the system, teachers are unlikely to be able to give children the breathing space they need to connect with their inner motivation to learn. What this thesis shows is that teachers need agency themselves in order to provide the conditions for agency in their students; this means they themselves need both support and space. It should be the role of policy makers to provide such conditions, rather than the punitive, high-stakes accountability regimes currently pressuring teachers to be, in turn, controlling towards their students.

10.4 Closing Remarks

This thesis has argued in favour of increasing children's autonomy and agency in their learning, and critiqued the way teacher control hinders this. However, as Sayer argues, alternative systems need to be viewed not only through the lens of the problem being criticised, but also by examining any problems the solutions would generate in turn. In the case of autonomy, the problem is that by relinquishing control to make space for autonomy, adults lose the ability to predict and guarantee exactly what children will learn in the form of learning objectives.

However, the issue may not be as intractable as it would appear. What this research and other child-centred pedagogies show is that autonomy and learning are not either/or, despite tensions. The teachers in our programme, despite their struggles, all provided pockets of space for children's autonomy. And whilst adhering strictly to learning objectives leaves little space for autonomy (of any conception), the deep, joyful kinds of learning our teachers and others have reported depend on autonomy and agency. Many of the strategies we offered teachers are in line with those advanced by Darling-Hammond et al. (2020) to support children's learning, well-being and holistic development. In particular, they emphasise the importance of meaningful work, as well as opportunities for inquiry supported by explicit instruction and scaffolding, and opportunities for children to question, explain and elaborate thoughts together. This means that teachers can engage with students in negotiations about the learning by taking a broader, more holistic view of learning and a longer-term, more process-oriented view of curricula: some aspects might be imposed or generated by the teacher, whilst giving children opportunities for agency and autonomy within the delimited space. Conversation, relationships, and the willingness to compromise might therefore be what is needed to resolve tensions between autonomy and adult objectives.

However, it takes a special kind of environment for teachers to be able to say, as one teacher in our programme, “You can't have [my lesson plan] because I don't know what the kids want to do yet.” Currently, the educational system in England makes it very difficult for schools to give teachers the flexibility they need, let alone the support and resources, to provide such conditions for autonomy and agency to flourish.

References

- Abuhamdeh, S., & Csikszentmihalyi, M. (2012a). The Importance of Challenge for the Enjoyment of Intrinsically Motivated, Goal-Directed Activities. *Personality and Social Psychology Bulletin*, 38(3), 317–330.
<https://doi.org/10.1177/0146167211427147>
- Abuhamdeh, S., & Csikszentmihalyi, M. (2012b). Attentional involvement and intrinsic motivation. *Motivation and Emotion*, 36(3), 257–267. <https://doi.org/10.1007/s11031-011-9252-7>
- Aelterman, N., Vansteenkiste, M., Van Keer, H., & Haerens, L. (2016). Changing teachers' beliefs regarding autonomy support and structure: The role of experienced psychological need satisfaction in teacher training. *Psychology of Sport and Exercise*, 23, 64–72. <https://doi.org/10.1016/j.psychsport.2015.10.007>
- Aguirre, J., & Speer, N. M. (1999). Examining the Relationship Between Beliefs and Goals in Teacher Practice. *The Journal of Mathematical Behavior*, 18(3), 327–356.
[https://doi.org/10.1016/S0732-3123\(99\)00034-6](https://doi.org/10.1016/S0732-3123(99)00034-6)
- Alberti, E. T., & Witryol, S. L. (1994). The Relationship Between Curiosity and Cognitive Ability in Third- and Fifth-Grade Children. *The Journal of Genetic Psychology*, 155(2), 129–145. <https://doi.org/10.1080/00221325.1994.9914767>
- Alexander, R. (2011). Evidence, rhetoric and collateral damage: The problematic pursuit of 'world class' standards. *Cambridge Journal of Education*, 41(3), 265–286.
<https://doi.org/10.1080/0305764X.2011.607153>
- Allan, N. P., Lonigan, C. J., & Wilson, S. B. (2013). Psychometric evaluation of the Children's Behavior Questionnaire-Very Short Form in preschool children using parent and teacher report. *Early Childhood Research Quarterly*, 28(2), 302–313.
<https://doi.org/10.1016/j.ecresq.2012.07.009>

- Almqvist, A.-L., & Almqvist, L. (2015). Making oneself heard – children’s experiences of empowerment in Swedish preschools. *Early Child Development and Care*, 185(4), 578–593. <https://doi.org/10.1080/03004430.2014.940931>
- Amabile, T. M. (1985). Motivation and creativity: Effects of motivational orientation on creative writers. *Journal of Personality and Social Psychology*, 48(2), 393.
- Anderman, E. M., & Maehr, M. L. (1994). Motivation and Schooling in the Middle Grades. *Review of Educational Research*, 64(2), 287–309.
- Andersen, M. M., Kiverstein, J., Miller, M., & Roepstorff, A. (2021). *Play in Predictive Minds: A Cognitive Theory of Play* [Preprint]. PsyArXiv. <https://doi.org/10.31234/osf.io/u86qy>
- Anmarkrud, Ø., & Bråten, I. (2009). Motivation for reading comprehension. *Learning and Individual Differences*, 19(2), 252–256. <https://doi.org/10.1016/j.lindif.2008.09.002>
- Archer, M. S. (1995). *Realist Social Theory: The Morphogenetic Approach*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511557675>
- Arnone, M. P., Grabowski, B. L., & Rynd, C. P. (1994). Curiosity as a personality variable influencing learning in a learner controlled lesson with and without advisement. *Educational Technology Research and Development*, 42(1), 5–20. <https://doi.org/10.1007/BF02298167>
- Assor, A., Kaplan, H., & Roth, G. (2002). Choice is good, but relevance is excellent: Autonomy-enhancing and suppressing teacher behaviours predicting students’ engagement in schoolwork. *British Journal of Educational Psychology*, 72(2), 261–278. <https://doi.org/10.1348/000709902158883>
- Assor, A., Roth, G., & Deci, E. L. (2004). The Emotional Costs of Parents’ Conditional Regard: A Self-Determination Theory Analysis. *Journal of Personality*, 72(1), 47–88. <https://doi.org/10.1111/j.0022-3506.2004.00256.x>

- Assor, A., Kaplan, H., Kanat-Maymon, Y., & Roth, G. (2005). Directly controlling teacher behaviors as predictors of poor motivation and engagement in girls and boys: The role of anger and anxiety. *Learning and Instruction, 15*(5), 397–413.
<https://doi.org/10.1016/j.learninstruc.2005.07.008>
- Baker, L., & Scher, D. (2002). Beginning Readers' Motivation for Reading in Relation to Parental Beliefs and Home Reading Experiences. *Reading Psychology, 23*(4), 239–269. <https://doi.org/10.1080/713775283>
- Baker, M., & Ryan, J. (2021). Playful provocations and playful mindsets: Teacher learning and identity shifts through playful participatory research. *International Journal of Play, 10*(1), 6–24. <https://doi.org/10.1080/21594937.2021.1878770>
- Baker, S. T., Le Courtois, S., & Eberhart, J. (2021). Making space for children's agency with playful learning. *International Journal of Early Years Education, 0*(0), 1–13.
<https://doi.org/10.1080/09669760.2021.1997726>
- Bakkenes, I., Vermunt, J. D., & Wubbels, T. (2010). Teacher learning in the context of educational innovation: Learning activities and learning outcomes of experienced teachers. *Learning and Instruction, 20*(6), 533–548.
<https://doi.org/10.1016/j.learninstruc.2009.09.001>
- Ball, S. J. (2003). The teacher's soul and the terrors of performativity. *Journal of Education Policy, 18*(2), 215–228. <https://doi.org/10.1080/0268093022000043065>
- Barandiaran, A., Muela, A., López de Arana, E., Larrea, I., & Vitoria, J. R. (2015). Exploratory behaviour, emotional wellbeing and childcare quality in preschool education. *Anales de Psicología, 31*(2), 570.
<https://doi.org/10.6018/analesps.31.2.171551>
- Baroody, A. E., & Diamond, K. E. (2012). Links Among Home Literacy Environment, Literacy Interest, and Emergent Literacy Skills in Preschoolers At Risk for Reading

- Difficulties. *Topics in Early Childhood Special Education*, 32(2), 78–87.
<https://doi.org/10.1177/0271121410392803>
- Baroody, A. E., & Diamond, K. E. (2013). Measures of preschool children's interest and engagement in literacy activities: Examining gender differences and construct dimensions. *Early Childhood Research Quarterly*, 28(2), 291–301.
<https://doi.org/10.1016/j.ecresq.2012.07.002>
- Baroody, A. E., & Diamond, K. E. (2016). Associations among preschool children's classroom literacy environment, interest and engagement in literacy activities, and early reading skills. *Journal of Early Childhood Research*, 14(2), 146–162.
<https://doi.org/10.1177/1476718X14529280>
- Baroody, A. E., Diamond, K. E., & Hong, S. (2006). *Children's interest measure*. Purdue University.
- Barton, A. C., & Tan, E. (2010). We Be Burnin'! Agency, Identity, and Science Learning. *Journal of the Learning Sciences*, 19(2), 187–229.
<https://doi.org/10.1080/10508400903530044>
- Baumeister, R. E., Bratslavsky, E., Muraven, M., & Tice, D. M. (1998). Ego Depletion: Is the Active Self a Limited Resource? *Journal of Personality and Social Psychology*, 74(5), 1252–1265.
- Baumfield, V. (2008). *Action research in the classroom*. Los Angeles, [Calif.].
- Bautista, A., Habib, M., Eng, A., & Bull, R. (2019). Purposeful play during learning centre time: From curriculum to practice. *Journal of Curriculum Studies*, 51(5), 715–736.
<https://doi.org/10.1080/00220272.2019.1611928>
- Beane, J. A. (1997). *Curriculum Integration: Designing the core of democratic education*. Teachers College Press.

- Becker, M., McElvany, N., & Kortenbruck, M. (2010). Intrinsic and extrinsic reading motivation as predictors of reading literacy: A longitudinal study. *Journal of Educational Psychology, 102*(4), 773–785. <https://doi.org/10.1037/a0020084>
- Begus, K., & Bonawitz, E. (2020). The rhythm of learning: Theta oscillations as an index of active learning in infancy. *Developmental Cognitive Neuroscience, 45*, 100810. <https://doi.org/10.1016/j.dcn.2020.100810>
- Benware, C. A., & Deci, E. L. (1984). Quality of Learning with an Active versus Passive Motivational Set. *American Educational Research Journal, 21*(4), 755. <https://doi.org/10.2307/1162999>
- Best, J. R., & Miller, P. H. (2010). A Developmental Perspective on Executive Function. *Child Development, 81*(6), 1641–1660. <https://doi.org/10.1111/j.1467-8624.2010.01499.x>
- Bevington, T. J. (2015). Appreciative evaluation of restorative approaches in schools. *Pastoral Care in Education, 33*(2), 105–115. <https://doi.org/10.1080/02643944.2015.1046475>
- Biesta, G. (2007). Why “What Works” Won’t Work: Evidence-Based Practice and the Democratic Deficit in Educational Research. *Educational Theory, 57*(1), 1–22. <https://doi.org/10.1111/j.1741-5446.2006.00241.x>
- Biesta, G. (2010). *Good education in an age of measurement ethics, politics, democracy / Gert J.J. Biesta*. Boulder, Colo.
- Biesta, G., & Tedder, M. (2007). Agency and learning in the lifecourse: Towards an ecological perspective. *Studies in the Education of Adults, 39*(2), 132–149. <https://doi.org/10.1080/02660830.2007.11661545>
- Birks, M., & Mills, J. (2015). *Grounded theory: A practical guide* (Second edition). SAGE.

- Blackburn, S. (2005). *Truth: A guide for the perplexed / Simon Blackburn*. London : Allen Lane, 2005.
- Bobis, J., Way, J., Anderson, J., & Martin, A. J. (2016). Challenging teacher beliefs about student engagement in mathematics. *Journal of Mathematics Teacher Education*, *19*(1), 33–55. <https://doi.org/10.1007/s10857-015-9300-4>
- Bodrova, E., & Leong, D. J. (2018). Tools of the Mind: The Vygotskian-Based Early Childhood Program. *Journal of Cognitive Education and Psychology*, *17*(3), 223–237. <https://doi.org/10.1891/1945-8959.17.3.223>
- Boekaerts, M. (2016). Engagement as an inherent aspect of the learning process. *Learning and Instruction*, *43*, 76–83. <https://doi.org/10.1016/j.learninstruc.2016.02.001>
- Boggiano, A. K., Main, D. S., & Katz, P. A. (1988). Children’s preference for challenge: The role of perceived competence and control. *Journal of Personality and Social Psychology*, *54*(1), 134–141.
- Bonawitz, E. (2010). The Double-edged Sword of Pedagogy: Instruction limits spontaneous exploration and discovery. *Cognition*.
- Bonett, D. G., & Wright, T. A. (2000). Sample size requirements for estimating pearson, kendall and spearman correlations. *Psychometrika*, *65*(1), 23–28. <https://doi.org/10.1007/BF02294183>
- Bowers, J. S. (2020). Reconsidering the Evidence That Systematic Phonics Is More Effective Than Alternative Methods of Reading Instruction. *Educational Psychology Review*. <https://doi.org/10.1007/s10648-019-09515-y>
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>

- Braun, V., & Clarke, V. (2019). Reflecting on reflexive thematic analysis. *Qualitative Research in Sport, Exercise and Health*, 11(4), 589–597.
<https://doi.org/10.1080/2159676X.2019.1628806>
- Braun, V., & Clarke, V. (2021a). *Can I use TA? Should I use TA? Should I not use TA? Comparing reflexive thematic analysis and other pattern-based qualitative analytic approaches*. 21, 37–47. <https://doi.org/10.1002/capr.12360>
- Braun, V., & Clarke, V. (2021b). Conceptual and design thinking for thematic analysis. *Qualitative Psychology*, Advance online publication.
<https://doi.org/10.1037/qup0000196>
- Bremner, N. (2019). From learner-centred to learning-centred: Becoming a ‘hybrid’ practitioner. *International Journal of Educational Research*, 97, 53–64.
<https://doi.org/10.1016/j.ijer.2019.06.012>
- British Education Research Association (BERA). (2011). *Ethical Guidelines for Educational Research*. <https://www.bera.ac.uk/wp-content/uploads/2014/02/BERA-Ethical-Guidelines-2011.pdf>
- Broadhead, P. (1992). Play-fighting, play or fighting? -- from parallel to co-operative play in the pre-school. *Early Years*, 13(1), 45–50. <https://doi.org/10.1080/0957514920130109>
- Brown, S. R. (1993). A primer on Q methodology. *Operant Subjectivity*, 16(3/4), 91–138.
- Browne, W. J., & Draper, D. (2000). Implementation and performance issues in the Bayesian and likelihood fitting of multilevel models. *Computational Statistics*, 15, 391–420.
- Burnett, C. (2017). Acknowledging and interrogating multiplicities: Towards a generous approach in evaluations of early literacy innovation and intervention. *Journal of Early Childhood Literacy*, 17(4), 522–550. <https://doi.org/10.1177/1468798416645851>

- Burnett, C., & Coldwell, M. (2020). Randomised controlled trials and the interventionisation of education. *Oxford Review of Education*, 0(0), 1–16.
<https://doi.org/10.1080/03054985.2020.1856060>
- Burrell, G., & Morgan, G. (1979). *Sociological Paradigms and Organisational Analysis: Elements of the Sociology of Corporate Life*. Routledge.
<https://doi.org/10.4324/9781315242804>
- Carbonneau, K. J., Orman, D. S. J. V., Lemberger-Truelove, M. E., & Atencio, D. J. (2020). Leveraging the Power of Observations: Locating the Sources of Error in the Individualized Classroom Assessment Scoring System. *Early Education and Development*, 31(1), 84–99. <https://doi.org/10.1080/10409289.2019.1617572>
- Carlton, M. P., & Winsler, A. (1998). Fostering intrinsic motivation in early childhood classrooms. *Early Childhood Education Journal*, 25(3), 159–166.
- Carter, C. P., Reschly, A. L., Lovelace, M. D., Appleton, J. J., & Thompson, D. (2012). Measuring student engagement among elementary students: Pilot of the Student Engagement Instrument—Elementary Version. *School Psychology Quarterly*, 27(2), 61. <https://doi.org/10.1037/a0029229>
- Cassidy, C., Christie, D., Coutts, N., Dunn, J., Sinclair, C., Skinner, D., & Wilson, A. (2008). Building communities of educational enquiry. *Oxford Review of Education*, 34(2), 217–235. <https://doi.org/10.1080/03054980701614945>
- Cerasoli, C. P., Nicklin, J. M., & Ford, M. T. (2014). Intrinsic motivation and extrinsic incentives jointly predict performance: A 40-year meta-analysis. *Psychological Bulletin*, 140(4), 980–1008. <https://doi.org/10.1037/a0035661>
- Chambers, C. T., & Johnston, C. (2002). Developmental Differences in Children’s Use of Rating Scales. *Journal of Pediatric Psychology*, 27(1), 27–36.
<https://doi.org/10.1093/jpepsy/27.1.27>

- Chan, M. C. E., & Clarke, D. (2017). Structured affordances in the use of open-ended tasks to facilitate collaborative problem solving. *ZDM*, *49*(6), 951–963.
<https://doi.org/10.1007/s11858-017-0876-2>
- Chaplin, W. F., John, O. P., & Goldberg, L. R. (1988). Conceptions of states and traits: Dimensional attributes with ideals as prototypes. *Journal of Personality and Social Psychology*, *54*(4), 541–557. <https://doi.org/10.1037/0022-3514.54.4.541>
- Cheon, S. H., Reeve, J., & Song, Y.-G. (2016). A Teacher-Focused Intervention to Decrease PE Students' Amotivation by Increasing Need Satisfaction and Decreasing Need Frustration. *Journal of Sport and Exercise Psychology*, *38*(3), 217–235.
<https://doi.org/10.1123/jsep.2015-0236>
- Chong, J. L., Lim, K. K., & Matchar, D. B. (2019). Population segmentation based on healthcare needs: A systematic review. *Systematic Reviews*, *8*(1), 202.
<https://doi.org/10.1186/s13643-019-1105-6>
- Chouinard, M. M., Harris, P. L., & Maratsos, M. P. (2007). Children's Questions: A Mechanism for Cognitive Development. *Monographs of the Society for Research in Child Development*, *72*(1), i–129.
- Chung, S., & Walsh, D. J. (2000). Unpacking child-centredness: A history of meanings. *Journal of Curriculum Studies*, *32*(2), 215–234.
<https://doi.org/10.1080/002202700182727>
- Ciani, K. D., Middleton, M. J., Summers, J. J., & Sheldon, K. M. (2010). Buffering against performance classroom goal structures: The importance of autonomy support and classroom community. *Contemporary Educational Psychology*, *35*(1), 88–99.
<https://doi.org/10.1016/j.cedpsych.2009.11.001>
- Clark, A. (2001). How to listen to very young children: The mosaic approach. *Child Care in Practice*, *7*(4), 333–341. <https://doi.org/10.1080/13575270108415344>

- Clark, A. (2005). Listening to and involving young children: A review of research and practice. *Early Child Development and Care*, 175(6), 489–505.
<https://doi.org/10.1080/03004430500131288>
- Clifford, M. M. (1991). Risk Taking: Theoretical, Empirical, and Educational Considerations. *Educational Psychologist*, 26(3–4), 263–297.
<https://doi.org/10.1080/00461520.1991.9653135>
- Coburn, C. E. (2001). Collective Sensemaking about Reading: How Teachers Mediate Reading Policy in Their Professional Communities. *Educational Evaluation and Policy Analysis*, 23(2), 145–170. <https://doi.org/10.3102/01623737023002145>
- Cremin, T., & Chappell, K. (2019). Creative pedagogies: A systematic review. *Research Papers in Education*, 0(0), 1–33. <https://doi.org/10.1080/02671522.2019.1677757>
- Csikszentmihalyi, M. (1990). *Flow: The psychology of optimal experience*. Harper Perennial.
- Culshaw, S. Y. (2019). *An exploration of what it means to be struggling as a secondary teacher in England*.
- Cuypers, S. E., & Haji, I. (2008). Educating for well-being and autonomy. *Theory and Research in Education*, 6(1), 71–93. <https://doi.org/10.1177/1477878507086731>
- d’Ailly, H. (2004). The Role of Choice in Children’s Learning: A Distinctive Cultural and Gender Difference in Efficacy, Interest, and Effort. *Canadian Journal of Behavioural Science / Revue Canadienne Des Sciences Du Comportement*, 36(1), 17–29.
<https://doi.org/10.1037/h0087212>
- Dahlberg, A., Ghaderi, A., Sarkadi, A., & Salari, R. (2019). SDQ in the Hands of Fathers and Preschool Teachers—Psychometric Properties in a Non-clinical Sample of 3–5-Year-Olds. *Child Psychiatry & Human Development*, 50(1), 132–141.
<https://doi.org/10.1007/s10578-018-0826-4>

- Darling-Hammond, L., Flook, L., Cook-Harvey, C., Barron, B., & Osher, D. (2020). Implications for educational practice of the science of learning and development. *Applied Developmental Science, 24*(2), 97–140. <https://doi.org/10.1080/10888691.2018.1537791>
- Darling-Hammond, L., & Richardson, N. (2009). Research Review / Teacher Learning: What Matters? *Educational Leadership, 66*, 46–53.
- de Kruijf, R. E. L., McWilliam, R. A., Ridley, S. M., & Wakely, M. B. (2000). Classification of teachers' interaction behaviors in early childhood classrooms. *Early Childhood Research Quarterly, 15*(2), 247–268. [https://doi.org/10.1016/S0885-2006\(00\)00051-X](https://doi.org/10.1016/S0885-2006(00)00051-X)
- de la Osa, N., Granero, R., Penelo, E., Domènech, J. M., & Ezpeleta, L. (2014). The Short and Very Short Forms of the Children's Behavior Questionnaire in a Community Sample of Preschoolers. *Assessment, 21*(4), 463–476. <https://doi.org/10.1177/1073191113508809>
- Deci, E. L. (1992). The Relation of Interest to the Motivation of Behavior: A Self-Determination Theory Perspective. In A. Renninger, S. Hidi, & A. Krapp, *The Role of Interest in Learning and Development*. Psychology Press.
- Deci, E. L., Eghrari, H., Patrick, B. C., & Leone, D. R. (1994). Facilitating internalization: The self-determination theory perspective. *Journal of Personality, 62*(1), 119–142.
- Deci, E. L., Koestner, R., & Ryan, R. M. (2001). Extrinsic Rewards and Intrinsic Motivation in Education: Reconsidered Once Again. *Review of Educational Research, 71*(1), 1–27. <https://doi.org/10.3102/00346543071001001>
- Desimone, L. M. (2009). Improving Impact Studies of Teachers' Professional Development: Toward Better Conceptualizations and Measures. *Educational Researcher, 38*(3), 181–199. <https://doi.org/10.3102/0013189X08331140>

- DfE. (2011). *Teachers' Standards. Guidance for school leaders, school staff and governing bodies*. Department for Education.
- DfE. (2019). *ITT Core Content Framework* (p. 49). Department for Education.
- DfE. (2021a). *Initial teacher training (ITT) market review report* (p. 56). Department for Education.
- DfE. (2021b). *Statutory framework for the early years foundation stage: Setting the standards for learning, development and care for children from birth to five* (p. 53). Department for Education.
- Diamond, A., & Ling, D. S. (2016). Conclusions about interventions, programs, and approaches for improving executive functions that appear justified and those that, despite much hype, do not. *Developmental Cognitive Neuroscience, 18*, 34–48.
<https://doi.org/10.1016/j.dcn.2015.11.005>
- Dockett, S., & Perry, B. (1999). Starting School: What Do the Children Say? *Early Child Development and Care, 159*(1), 107–119. <https://doi.org/10.1080/0300443991590109>
- Doctoroff, G. L., Fisher, P. H., Burrows, B. M., & Edman, M. T. (2016). Preschool Children's Interest, Social-emotional Skills, and Emergent Mathematical Skills. *Psychology in the Schools, 53*(4), 390–403. <https://doi.org/10.1002/pits.21912>
- Downer, J. T., Booren, L. M., Lima, O. K., Luckner, A. E., & Pianta, R. C. (2010). The Individualized Classroom Assessment Scoring System (inCLASS): Preliminary Reliability and Validity of a System for Observing Preschoolers' Competence in Classroom Interactions. *Early Childhood Research Quarterly, 25*(1), 1–16.
<https://doi.org/10.1016/j.ecresq.2009.08.004>
- Downer, J. T., Rimm-Kaufman, S. E., & Pianta, R. C. (2007). How Do Classroom Conditions and Children's Risk for School Problems Contribute to Children's Behavioral Engagement in Learning? *School Psychology Review, 36*(3), 413–432.

- Duffy, G., & Elwood, J. (2013). The perspectives of 'disengaged' students in the 14-19 phase on motivations and barriers to learning within the contexts of institutions and classrooms. *London Review of Education*, 11(2), 112–126.
<https://doi.org/10.1080/14748460.2013.799808>
- Dunst, C. J., Jones, T., Johnson, M., Raab, M., & Hamby, D. W. (2011). Role of children's interests in early literacy and language development. *Center for Early Literacy Learning*, 4(5).
http://www.earlyliteracylearning.org/cellreviews/cellreviews_v4_n5.pdf
- Early Excellence. (2021, January 18). What is Continuous Provision and why is it important? *Early Excellence*. <https://earlyexcellence.com/latest-news/press-articles/using-continuous-provision/>
- Eberhart, J. (2020). *Young Children's Executive Functions in Context: Classroom Experiences and Measurement Approaches*. University of Cambridge.
- Eccles, J. S. (2016). Engagement: Where to next? *Learning and Instruction*, 43, 71–75.
<https://doi.org/10.1016/j.learninstruc.2016.02.003>
- Edwards, A. (2005). Relational agency: Learning to be a resourceful practitioner. *International Journal of Educational Research*, 43(3), 168–182.
<https://doi.org/10.1016/j.ijer.2006.06.010>
- EIF. (2015). *Translating the evidence. A brief guide to the Early Intervention Foundation's procedures for identifying, assessing and disseminating information about early intervention programmes and their evidence*. Early Intervention Foundation.
<http://www.eif.org.uk/wp-content/uploads/2015/09/08-09-15-TRANSLATING-THE-EVIDENCE-IPR-Review.pdf>

- Einarsdottir, J. (2005). Playschool in pictures: Children's photographs as a research method. *Early Child Development and Care*, 175(6), 523–541.
<https://doi.org/10.1080/03004430500131320>
- Elen, J., Clarebout, G., Léonard, R., & Lowyck, J. (2007). Student-centred and teacher-centred learning environments: What students think. *Teaching in Higher Education*, 12(1), 105–117. <https://doi.org/10.1080/13562510601102339>
- Ellaway, R. H., Pusic, M., Yavner, S., & Kalet, A. L. (2014). Context matters: Emergent variability in an effectiveness trial of online teaching modules. *Medical Education*, 48(4), 386–396. <https://doi.org/10.1111/medu.12389>
- Elliot, A. J., & Church, M. A. (1997). A hierarchical model of approach and avoidance achievement motivation. *Journal of Personality and Social Psychology*, 72(1), 218.
- Ellis, S., & Moss, G. (2014). Ethics, education policy and research: The phonics question reconsidered. *British Educational Research Journal*, 40(2), 241–260.
<https://doi.org/10.1002/berj.3039>
- Emirbayer, M., & Mische, A. (1998). What Is Agency? *American Journal of Sociology*, 103(4), 962–1023. <https://doi.org/10.1086/231294>
- Engel, S. (2009). Is Curiosity Vanishing? *Journal of the American Academy of Child & Adolescent Psychiatry*, 48(8), 777–779.
<https://doi.org/10.1097/CHI.0b013e3181aa03b0>
- Engel, S. (2011). Children's Need to Know: Curiosity in schools. *Harvard Educational Review*, 81(4), 625–645.
- Ephgrave, A. (2018). *Planning in the moment with young children: A practical guide for early years practitioners and parents* / [by Anna Ephgrave]. Milton Park, Abingdon, Oxon : Routledge, 2018.

- Eshach, H., Dor-Ziderman, Y., & Yefroimsky, Y. (2014). Question Asking in the Science Classroom: Teacher Attitudes and Practices. *Journal of Science Education and Technology*, 23(1), 67–81. <https://doi.org/10.1007/s10956-013-9451-y>
- Fielding, M. (2001). OFSTED, Inspection and the Betrayal of Democracy. *Journal of Philosophy of Education*, 35(4), 695–709. <https://doi.org/10.1111/1467-9752.00254>
- Finn, J. D., & Rock, D. A. (1997). Academic success among students at risk for school failure. *Journal of Applied Psychology*, 82(2), 221.
- Fisher, H. (2011). Inside The Primary Classroom: Examples of Dissatisfaction Behind A Veil of Compliance. *British Journal of Educational Studies*, 59(2), 121–141. <https://doi.org/10.1080/00071005.2011.567969>
- Fisher, J. (2016). *Interacting or Interfering? Improving Interactions in the Early Years*. Open University Press.
- Fisher, K. R., Hirsh-Pasek, K., Newcombe, N., & Golinkoff, R. M. (2013). Taking Shape: Supporting Preschoolers' Acquisition of Geometric Knowledge Through Guided Play. *Child Development*, 84(6), 1872–1878. <https://doi.org/10.1111/cdev.12091>
- Flink, C., Boggiano, A. K., & Barrett, M. (1990). Controlling teaching strategies: Undermining children's self-determination and performance. *Journal of Personality and Social Psychology*, 59(5), 916–924. <https://doi.org/10.1037/0022-3514.59.5.916>
- Flint, A. S., Zisook, K., & Fisher, T. R. (2011). Not a one-shot deal: Generative professional development among experienced teachers. *Teaching and Teacher Education*, 27(8), 1163–1169. <https://doi.org/10.1016/j.tate.2011.05.009>
- Flowerday, T., Schraw, G., & Stevens, J. (2004). The Role of Choice and Interest in Reader Engagement. *The Journal of Experimental Education*, 72(2), 93–114.

- Flowerday, T., & Shell, D. F. (2015). Disentangling the effects of interest and choice on learning, engagement, and attitude. *Learning and Individual Differences, 40*, 134–140. <https://doi.org/10.1016/j.lindif.2015.05.003>
- Flunger, B., Pretsch, J., Schmitt, M., & Ludwig, P. (2013). The role of explicit need strength for emotions during learning. *Learning and Individual Differences, 23*, 241–248. <https://doi.org/10.1016/j.lindif.2012.10.001>
- Forman, D. R. (2007). Autonomy, compliance, and internalization. In *Socioemotional development in the toddler years: Transitions and transformations* (pp. 285–319). The Guilford Press.
- Fortier, M. S., Vallerand, R. J., & Guay, F. (1995). Academic Motivation and School Performance: Toward a Structural Model. *Contemporary Educational Psychology, 20*(3), 257–274. <https://doi.org/10.1006/ceps.1995.1017>
- Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research, 74*(1), 59–109.
- Fredricks, J. A., Filsecker, M., & Lawson, M. A. (2016). Student engagement, context, and adjustment: Addressing definitional, measurement, and methodological issues. *Learning and Instruction, 43*, 1–4. <https://doi.org/10.1016/j.learninstruc.2016.02.002>
- Fredricks, J. A., Wang, M.-T., Schall Linn, J., Hofkens, T. L., Sung, H., Parr, A., & Allerton, J. (2016). Using qualitative methods to develop a survey measure of math and science engagement. *Learning and Instruction, 43*, 5–15. <https://doi.org/10.1016/j.learninstruc.2016.01.009>
- Freedman-Doan, C., Wigfield, A., Eccles, J. S., Blumenfeld, P., Arbreton, A., & Harold, R. D. (2000). What Am I Best At? Grade and Gender Differences in Children's Beliefs About Ability Improvement. *Journal of Applied Developmental Psychology, 21*(4), 379–402. [https://doi.org/10.1016/S0193-3973\(00\)00046-0](https://doi.org/10.1016/S0193-3973(00)00046-0)

- Freelon, D. (2013). ReCal OIR: Ordinal, Interval, and Ratio Intercoder Reliability as a Web Service. *International Journal of Internet Science*, 8(1), 10–16.
- Freeman, M. (1998). The Right to Be Heard. *Adoption & Fostering*, 22(4), 50–59.
<https://doi.org/10.1177/030857599802200408>
- Freire, P. (1972/ 2000). *Pedagogy of the oppressed* (30th Anniversary Edition). Continuum.
- Frierson, P. R. (2016). Making Room for Children’s Autonomy: Maria Montessori’s Case for Seeing Children’s Incapacity for Autonomy as an External Failing. *Journal of Philosophy of Education*, 50(3), 332–350. <https://doi.org/10.1111/1467-9752.12134>
- Fuhs, M. W., Farran, D. C., & Nesbitt, K. T. (2013). Preschool classroom processes as predictors of children’s cognitive self-regulation skills development. *School Psychology Quarterly*, 28(4), 347–359. <https://doi.org/10.1037/spq0000031>
- Fulmer, S. M., D’Mello, S. K., Strain, A., & Graesser, A. C. (2015). Interest-based text preference moderates the effect of text difficulty on engagement and learning. *Contemporary Educational Psychology*, 41, 98–110.
<https://doi.org/10.1016/j.cedpsych.2014.12.005>
- Fulmer, S. M., & Tulis, M. (2013). Changes in interest and affect during a difficult reading task: Relationships with perceived difficulty and reading fluency. *Learning and Instruction*, 27, 11–20. <https://doi.org/10.1016/j.learninstruc.2013.02.001>
- Glaser, B., & Strauss, A. (1967). *The Discovery of Grounded Theory: Strategies for Qualitative Research*.
- Godwin, K. E., Almeda, Ma. V., Seltman, H., Kai, S., Skerbetz, M. D., Baker, R. S., & Fisher, A. V. (2016). Off-task behavior in elementary school children. *Learning and Instruction*, 44, 128–143. <https://doi.org/10.1016/j.learninstruc.2016.04.003>
- Goldacre, B. (2013). *Building evidence into education*.
http://dera.ioe.ac.uk/17530/7/ben%20goldacre%20paper_Redacted.pdf

- Goodman, R. (1997). The Strengths and Difficulties Questionnaire: A Research Note. *Journal of Child Psychology and Psychiatry*, 38(5), 581–586.
<https://doi.org/10.1111/j.1469-7610.1997.tb01545.x>
- Goouch, K. (2008). Understanding playful pedagogies, play narratives and play spaces. *Early Years*, 28(1), 93–102. <https://doi.org/10.1080/09575140701815136>
- Gopnik, A. (2012). Scientific Thinking in Young Children: Theoretical Advances, Empirical Research, and Policy Implications. *Science*, 337(6102), 1623–1627.
<https://doi.org/10.1126/science.1223416>
- Gopnik, A. (2016). *The Gardener and the Carpenter: What the New Science of Child Development Tells us about the Relationship Between Parents and Children*. The Bodley Head.
- Gore, J. M., & Gitlin, A. D. (2004). [RE]Visioning the academic–teacher divide: Power and knowledge in the educational community. *Teachers and Teaching*, 10(1), 35–58.
<https://doi.org/10.1080/13540600320000170918>
- Gottfried, A. E. (1985). Academic intrinsic motivation in elementary and junior high school students. *Journal of Educational Psychology*, 77(6), 631.
- Gottfried, A. E., Fleming, J. S., & Gottfried, A. W. (2001). Continuity of academic intrinsic motivation from childhood through late adolescence: A longitudinal study. *Journal of Educational Psychology*, 93(1), 3–13. <https://doi.org/10.1037//0022-0663.93.1.3>
- Gottfried, A. E., Marcoulides, G. A., Gottfried, A. W., & Oliver, P. H. (2013). Longitudinal Pathways From Math Intrinsic Motivation and Achievement to Math Course Accomplishments and Educational Attainment. *Journal of Research on Educational Effectiveness*, 6(1), 68–92. <https://doi.org/10.1080/19345747.2012.698376>
- Gottfried, A. E., Preston, K. S. J., Gottfried, A. W., Oliver, P. H., Delany, D. E., & Ibrahim, S. M. (2016). Pathways from parental stimulation of children’s curiosity to high

- school science course accomplishments and science career interest and skill. *International Journal of Science Education*, 38(12), 1972–1995.
<https://doi.org/10.1080/09500693.2016.1220690>
- Graves, S. B., & Larkin, E. (2006). Lessons from Erikson. *Journal of Intergenerational Relationships*, 4(2), 61–71. https://doi.org/10.1300/J194v04n02_05
- Gray, P. (2017). Self-Directed Education—Unschooling and Democratic Schooling. In P. Gray, *Oxford Research Encyclopedia of Education*. Oxford University Press.
<https://doi.org/10.1093/acrefore/9780190264093.013.80>
- Greany, T., & Waterhouse, J. (2016). Rebels against the system: Leadership agency and curriculum innovation in the context of school autonomy and accountability in England. *International Journal of Educational Management*, 30(7), 1188–1206.
<https://doi.org/10.1108/IJEM-11-2015-0148>
- Green, C. (2013). A Sense of Autonomy in Young Children’s Special Places. *International Journal of Early Childhood Environmental Education*, 1(1), 8–31.
- Green, L. W., & Glasgow, R. E. (2006). Evaluating the Relevance, Generalization, and Applicability of Research: Issues in External Validation and Translation Methodology. *Evaluation & the Health Professions*, 29(1), 126–153.
<https://doi.org/10.1177/0163278705284445>
- Gregoire, M. (2003). Is It a Challenge or a Threat? A Dual-Process Model of Teachers’ Cognition and Appraisal Processes During Conceptual Change. *Educational Psychology Review*, 33.
- Gripton, C., & Knight, R. (2020). Walking the Talk: Moving forwards with sustained shared thinking and dialogic teaching. *FORUM*, 62(1), 31.
<https://doi.org/10.15730/forum.2020.62.1.31>

- Grolnick, W. S., & Ryan, R. M. (1987). Autonomy in children's learning: An experimental and individual difference investigation. *Journal of Personality and Social Psychology*, 52(5), 890–898. <https://doi.org/10.1037//0022-3514.52.5.890>
- Gruber, M. J., Gelman, B. D., & Ranganath, C. (2014). States of Curiosity Modulate Hippocampus-Dependent Learning via the Dopaminergic Circuit. *Neuron*, 84(2), 486–496. <https://doi.org/10.1016/j.neuron.2014.08.060>
- Guay, F., Ratelle, C. F., Roy, A., & Litalien, D. (2010). Academic self-concept, autonomous academic motivation, and academic achievement: Mediating and additive effects. *Learning and Individual Differences*, 20(6), 644–653. <https://doi.org/10.1016/j.lindif.2010.08.001>
- Guay, F., Roy, A., & Valois, P. (2017). Teacher structure as a predictor of students' perceived competence and autonomous motivation: The moderating role of differentiated instruction. *British Journal of Educational Psychology*, 87(2), 224–240. <https://doi.org/10.1111/bjep.12146>
- Guthrie, J. T., Hoa, A. L. W., Wigfield, A., Tonks, S. M., Humenick, N. M., & Littles, E. (2007). Reading motivation and reading comprehension growth in the later elementary years. *Contemporary Educational Psychology*, 32(3), 282–313. <https://doi.org/10.1016/j.cedpsych.2006.05.004>
- Hall, C. L., Guo, B., Valentine, A. Z., Groom, M. J., Daley, D., Sayal, K., & Hollis, C. (2019). The validity of the Strengths and Difficulties Questionnaire (SDQ) for children with ADHD symptoms. *PLOS ONE*, 14(6), e0218518. <https://doi.org/10.1371/journal.pone.0218518>
- Hantzopoulos, M. (2013). The Fairness Committee: Restorative Justice in a Small Urban Public High School. *Prevention Researcher*, 20(1), 7–10.

- Harackiewicz, J. M., Barron, K. E., Carter, S. M., Lehto, A. T., & Elliot, A. J. (1997). Predictors and consequences of achievement goals in the college classroom: Maintaining interest and making the grade. *Journal of Personality and Social Psychology*, 73(6), 1284–1295. <https://doi.org/10.1037/0022-3514.73.6.1284>
- Hargreaves, D. H. (1996). *Teaching as a research-based profession: Possibilities and prospects: Vol. Teacher Training Agency annual lecture*. Teacher Training Agency.
- Hargreaves, E., Berry, R., Lai, Y. C., Leung, P., Scott, D., & Stobart, G. (2013). Teachers' experiences of autonomy in Continuing Professional Development: Teacher Learning Communities in London and Hong Kong. *Teacher Development*, 17(1), 19–34. <https://doi.org/10.1080/13664530.2012.748686>
- Harris Mackay, S. (2021). *Story Workshop: New Possibilities for Young Writers*. Heinemann.
- Harrison, L. (2007). From Authoritarian to Restorative Schools. *Reclaiming Children and Youth: The Journal of Strength-Based Interventions*, 16(2), 17–20.
- Harter, S. (1978). Pleasure Derived from Challenge and the Effects of Receiving Grades on Children's Difficulty Level Choices. *Child Development*, 49(3), 788–799. <https://doi.org/10.2307/1128249>
- Hauge, K. (2019). Teachers' collective professional development in school: A review study. *Cogent Education*, 6(1), 1619223. <https://doi.org/10.1080/2331186X.2019.1619223>
- Hayes, A. F., & Krippendorff, K. (2007). Answering the Call for a Standard Reliability Measure for Coding Data. *Communication Methods and Measures*, 1(1), 77–89. <https://doi.org/10.1080/19312450709336664>
- Hedges, H. (2021). Contemporary principles to lead understandings of children's learning: Synthesizing Vygotsky, Rogoff, Wells and Lindfors. *Early Child Development and Care*, 191(7–8), 1056–1065. <https://doi.org/10.1080/03004430.2020.1849169>

- Heemskerk, C. H. H. M., & Malmberg, L.-E. (2020). Students' observed engagement in lessons, instructional activities, and learning experiences. *Frontline Learning Research*, 8(6), 38–58. <https://doi.org/10.14786/flr.v8i6.613>
- Hektner, J. M., Schmidt, J. A., & Csikszentmihalyi, M. (2006). *Experience sampling method: Measuring the quality of everyday life*. Sage.
- Hillage, J. (1998). *Excellence in research on schools*. Dept. for Education and Employment.
- Hoddinott, P. (2015). A new era for intervention development studies. *Pilot and Feasibility Studies*, 1(1), 36. <https://doi.org/10.1186/s40814-015-0032-0>
- Hofmann, R., & Rainio, A. P. (2007). 'It doesn't matter what part you play, it just matters that you're there.' Towards shared agency in narrative play activity in school. In R. Alanen & S. Pöyhönen (Eds.), *Language in action. Vygotsky and Leontievian legacy today* (pp. 308–328). Cambridge Scholars Publishing.
- Hogan, R., Johnson, J., & Briggs, S. R. (Eds.). (1997). *Handbook of Personality Psychology*. Academic Press. <https://www.dawsonera.com:443/abstract/9780080533179>
- Holland, P. (2000). Take the Toys from the Boys? An Examination of the Genesis of Policy and the Appropriateness of Adult Perspectives in the Area of War, Weapon and Superhero Play. *Citizenship, Social and Economics Education*, 4(2), 92–108. <https://doi.org/10.2304/csee.2000.4.2.92>
- hooks, bell. (1994). *Teaching to transgress: Education as the practice of freedom*. Routledge.
- Hospel, V., & Galand, B. (2016). Are both classroom autonomy support and structure equally important for students' engagement? A multilevel analysis. *Learning and Instruction*, 41, 1–10. <https://doi.org/10.1016/j.learninstruc.2015.09.001>

- Howard, J. (2002). Eliciting Young Children's Perceptions of Play, Work and Learning Using the Activity Apperception Story Procedure. *Early Child Development and Care*, 172(5), 489–502. <https://doi.org/10.1080/03004430214548>
- Hox, J. J., Moerbeek, M., & van de Schoot, R. (2010). *Multilevel Analysis: Techniques and Applications, Second Edition*. Taylor & Francis Group.
<http://ebookcentral.proquest.com/lib/cam/detail.action?docID=574571>
- Hughes, C., Lindberg, A., & Devine, R. T. (2018). Autonomy Support in Toddlerhood: Similarities and Contrasts Between Mothers and Fathers. *Journal of Family Psychology*, 32(7), 915–925. <https://doi.org/10.1037/fam0000450>
- Hughes, J. N., Luo, W., Kwok, O.-M., & Loyd, L. K. (2008). Teacher–Student Support, Effortful Engagement, and Achievement: A 3-Year Longitudinal Study. *Journal of Educational Psychology*, 100(1), 1–14. <https://doi.org/10.1037/0022-0663.100.1.1>
- Hume, L. E., Lonigan, C. J., & McQueen, J. D. (2015). Children's literacy interest and its relation to parents' literacy-promoting practices. *Journal of Research in Reading*, 38(2), 172–193. <https://doi.org/10.1111/j.1467-9817.2012.01548.x>
- Hutchings, M. (2015). *Exam factories? The impact of accountability measures on children and young people*. National Union of Teachers.
- Inoue, N. (2007). Why face a challenge?: The reason behind intrinsically motivated students' spontaneous choice of challenging tasks. *Learning and Individual Differences*, 17(3), 251–259. <https://doi.org/10.1016/j.lindif.2007.02.002>
- Ishimine, K., & Tayler, C. (2014). Assessing Quality in Early Childhood Education and Care. *European Journal of Education*, 49(2), 272–290. <https://doi.org/10.1111/ejed.12043>
- Iyengar, S. S., & Lepper, M. R. (1999). Rethinking the value of choice: A cultural perspective on intrinsic motivation. *Journal of Personality and Social Psychology*, 76(3), 349–366. <https://doi.org/10.1037/0022-3514.76.3.349>

- Iyengar, S. S., & Lepper, M. R. (2000). When choice is demotivating: Can one desire too much of a good thing? *Journal of Personality and Social Psychology*, 79(6), 995–1006. <https://doi.org/10.1037/0022-3514.79.6.995>
- Jang, H., Reeve, J., & Deci, E. L. (2010). Engaging students in learning activities: It is not autonomy support or structure but autonomy support and structure. *Journal of Educational Psychology*, 102(3), 588.
- Jang, H., Reeve, J., & Halusic, M. (2016). A New Autonomy-Supportive Way of Teaching That Increases Conceptual Learning: Teaching in Students' Preferred Ways. *The Journal of Experimental Education*, 84(4), 686–701. <https://doi.org/10.1080/00220973.2015.1083522>
- Jarvis, P. (2018). Simplistic Beginnings? The Invisibility of Sustained Shared Thinking in Ofsted Advice Documents. *FORUM*, 60(3), 301312. <https://doi.org/10.15730/forum.2018.60.3.301312>
- Jensen, H., Pyle, A., Zosh, J. M., Ebrahim, H. B., Scherman, A. Z., Reunamo, J., & Hamre, B. K. (2019). *Play facilitation: The science behind the art of engaging young children*. <https://doi.org/10.13140/RG.2.2.24373.12004>
- Jirout, J., & Klahr, D. (2012). Children's scientific curiosity: In search of an operational definition of an elusive concept. *Developmental Review*, 32(2), 125–160. <https://doi.org/10.1016/j.dr.2012.04.002>
- Jirout, J., & Klahr, D. (2020). Questions – And Some Answers – About Young Children's Questions. *Journal of Cognition and Development*, 21(5), 729–753. <https://doi.org/10.1080/15248372.2020.1832492>
- Jones, S. H., Johnson, M. L., & Campbell, B. D. (2015). Hot factors for a cold topic: Examining the role of task-value, attention allocation, and engagement on conceptual

change. *Contemporary Educational Psychology*, 42, 62–70.

<https://doi.org/10.1016/j.cedpsych.2015.04.004>

Kain, D. L. (2003). Owing Significance: The Critical Incident Technique in Research. In *Foundations for Research*. Routledge.

Kang, M. J., Hsu, M., Krajbich, I. M., Loewenstein, G., McClure, S. M., Wang, J. T., & Camerer, C. F. (2009). The Wick in the Candle of Learning: Epistemic Curiosity Activates Reward Circuitry and Enhances Memory. *Psychological Science*, 20(8), 963–973. <https://doi.org/10.1111/j.1467-9280.2009.02402.x>

Kashdan, T. B., & Steger, M. F. (2007). Curiosity and pathways to well-being and meaning in life: Traits, states, and everyday behaviors. *Motivation and Emotion*, 31(3), 159–173. <https://doi.org/10.1007/s11031-007-9068-7>

Katz, I., & Assor, A. (2006). When Choice Motivates and When It Does Not. *Educational Psychology Review*, 19(4), 429. <https://doi.org/10.1007/s10648-006-9027-y>

Kim, B. S., & Darling, L. F. (2009). Monet, Malaguzzi, and the Constructive Conversations of Preschoolers in a Reggio-Inspired Classroom. *Early Childhood Education Journal*, 37(2), 137–145. <https://doi.org/10.1007/s10643-009-0323-2>

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.

Kittredge, A. K., Day, N., Janik-Blaskova, L., Zhang, H., & Baker, S. (2018). Is Child-Led Science Learning Meaningful and Feasible? Developing a Theory of Change with Teacher Co-Researchers. *Tailoring Pedagogy to Suit the Learner: Examining the Impacts of Different Pedagogical Approaches*. American Educational Research Association Annual Meeting, New York, N.Y.

- Klahr, D., & Nigam, M. (2004). The equivalence of learning paths in early science instruction effects of direct instruction and discovery learning. *Psychological Science*, *15*(10), 661–667.
- Korsgaard, M. T. (2020). Exemplarity and education: Retuning educational research. *British Educational Research Journal*, *46*(6), 1357–1370. <https://doi.org/10.1002/berj.3636>
- Korthagen, F. (2017). Inconvenient truths about teacher learning: Towards professional development 3.0. *Teachers and Teaching: Theory and Practice*, *23*(4), 387–405. Scopus. <https://doi.org/10.1080/13540602.2016.1211523>
- Krippendorff, K. (2011). Agreement and Information in the Reliability of Coding. *Communication Methods and Measures*, *5*(2), 93–112. <https://doi.org/10.1080/19312458.2011.568376>
- Kuhn, D., & Ho, V. (1980). Self-directed activity and cognitive development. *Journal of Applied Developmental Psychology*, *1*(2). [https://doi.org/10.1016/0193-3973\(80\)90003-9](https://doi.org/10.1016/0193-3973(80)90003-9)
- Kumu Relationship Mapping Software*. (2020). <https://kumu.io>
- Laevers, F. (1994). *The Leuven Involvement Scale for Young Children, LIS-YC: Manual and video tape*. Centre for Experiential Education.
- Laevers, F. (2000). Forward to Basics! Deep-Level-Learning and the Experiential Approach. *Early Years*, *20*(2), 20–29. <https://doi.org/10.1080/0957514000200203>
- Laevers, F. (2003). Experiential Education: Making Care and Education More Effective Through Well-Being and Involvement. In F. Laevers & L. Haylen (Eds.), *Involvement of Children and Teacher Style: Insights from an International Study of Experiential Education* (Vol. 35). Leuven University Press.

- Laevers, F., & Declercq, B. (2018). How well-being and involvement fit into the commitment to children's rights. *European Journal of Education*, 53(3), 325–335.
<https://doi.org/10.1111/ejed.12286>
- Laevers, F., Declercq, B., & Jackaman, S. (2011) *Observing Involvement in Children from 6 to 12 years: A DVD Training Pack*. Centre for Experiential Education [Leuven] and Kent County Council. CEGO, Leuven.
- Laevers, F., Declercq, B., Marin, C., Moons, J. & Stanton, F. (2010) *Observing Involvement in Children from birth to 6 years: A DVD Training Pack*. Centre for Experiential Education [Leuven] and Kent County Council. CEGO, Leuven.
- Langhout, R. D. (2005). Acts of Resistance: Student (In)visibility. *Culture & Psychology*, 11(2), 123–158. <https://doi.org/10.1177/1354067X05052348>
- Lauriola, M., Litman, J. A., Mussel, P., De Santis, R., Crowson, H. M., & Hoffman, R. R. (2015). Epistemic curiosity and self-regulation. *Personality and Individual Differences*, 83, 202–207. <https://doi.org/10.1016/j.paid.2015.04.017>
- Lave, J., & Wenger, E. (1991). *Situated learning. Legitimate peripheral participation*. Cambridge University Press.
- Law, J. (2004). *After method: Mess in social science research / John Law*. (1st ed.). London.
- Le Courtois, S. (2017). 'Love of learning' as a family of constructs: *The evidence from a study with primary school children using measures of intrinsic motivation, mastery orientation and interest value*. University of Cambridge.
- Le Courtois, S. (2019). On Judgement. *FERSA University of Cambridge Blog*.
<https://fersacambridge.com/2019/10/07/on-judgement/>
- Lee, E., & Hannafin, M. J. (2016). A design framework for enhancing engagement in student-centered learning: Own it, learn it, and share it. *Educational Technology*

Research and Development, 64(4), 707–734. <https://doi.org/10.1007/s11423-015-9422-5>

Lee, M. K., Baker, S., & Whitebread, D. (2018). Culture-specific links between maternal executive function, parenting, and preschool children's executive function in South Korea. *British Journal of Educational Psychology*, 88(2), 216–235.

<https://doi.org/10.1111/bjep.12221>

Lemos, M. S., & Veríssimo, L. (2014). The Relationships between Intrinsic Motivation, Extrinsic Motivation, and Achievement, Along Elementary School. *Procedia - Social and Behavioral Sciences*, 112, 930–938.

<https://doi.org/10.1016/j.sbspro.2014.01.1251>

Lemov, D. (2012). *Teach like a champion field guide: A practical resource to make the 49 techniques your own / Doug Lemov*. San Francisco, Calif. : Jossey-Bass, c2012.

Lepper, M. R., Corpus, J. H., & Iyengar, S. S. (2005). Intrinsic and Extrinsic Motivational Orientations in the Classroom: Age Differences and Academic Correlates. *Journal of Educational Psychology*, 97(2), 184–196. <https://doi.org/10.1037/0022-0663.97.2.184>

Lepper, M. R., & Greene, D. (1975). Turning play into work: Effects of adult surveillance and extrinsic rewards on children's intrinsic motivation. *Journal of Personality and Social Psychology*, 31(3), 479–486. <https://doi.org/10.1037/h0076484>

Levesque, C., Zuehlke, A. N., Stanek, L. R., & Ryan, R. M. (2004). Autonomy and Competence in German and American University Students: A Comparative Study Based on Self-Determination Theory. *Journal of Educational Psychology*, 96(1), 68–84. <https://doi.org/10.1037/0022-0663.96.1.68>

Lewis, C., Perry, R., & Murata, A. (2006). How should research contribute to instructional improvement? The case of lesson study. *Educational Researcher*, 35(3), 3–14.

- Lietaert, S., Roorda, D., Laevers, F., Verschueren, K., & De Fraine, B. (2015). The gender gap in student engagement: The role of teachers' autonomy support, structure, and involvement. *British Journal of Educational Psychology*, 85(4), 498–518.
<https://doi.org/10.1111/bjep.12095>
- Lillard, A. S. (2013). Playful learning and Montessori education. *American Journal of Play*, 5(2), 157.
- Lillard, A. S. (2017). *Montessori: The science behind the genius / Angeline Stoll Lillard* (Third edition.). New York : Oxford University Press.
- Lillard, A. S. (2019). Shunned and Admired: Montessori, Self-Determination, and a Case for Radical School Reform. *Educational Psychology Review*, 31(4), 939–965.
<https://doi.org/10.1007/s10648-019-09483-3>
- Lipponen, L., & Kumpulainen, K. (2011). Acting as accountable authors: Creating interactional spaces for agency work in teacher education. *Teaching and Teacher Education*, 27(5), 812–819. <https://doi.org/10.1016/j.tate.2011.01.001>
- Lipponen, L., Rajala, A., Hilppö, J., & Paananen, M. (2016). Exploring the foundations of visual methods used in research with children. *European Early Childhood Education Research Journal*, 24(6), 936–946. <https://doi.org/10.1080/1350293X.2015.1062663>
- Litman, J. (2005). Curiosity and the pleasures of learning: Wanting and liking new information. *Cognition & Emotion*, 19(6), 793–814.
<https://doi.org/10.1080/02699930541000101>
- Litman, J. A. (2008). Interest and deprivation factors of epistemic curiosity. *Personality and Individual Differences*, 44(7), 1585–1595. <https://doi.org/10.1016/j.paid.2008.01.014>
- Lofthouse, R. (2019). Coaching in education: A professional development process in formation. *Professional Development in Education*, 45(1), 33–45.
<https://doi.org/10.1080/19415257.2018.1529611>

- Lofthouse, R., & Thomas, U. (2017). Concerning collaboration: Teachers' perspectives on working in partnerships to develop teaching practices. *Professional Development in Education*, 43(1), 36–56. <https://doi.org/10.1080/19415257.2015.1053570>
- Lomax, H. (2012). Contested voices? Methodological tensions in creative visual research with children. *International Journal of Social Research Methodology*, 15(2), 105–117. <https://doi.org/10.1080/13645579.2012.649408>
- Lortie-Forgues, H., & Inglis, M. (2019). Rigorous Large-Scale Educational RCTs Are Often Uninformative: Should We Be Concerned? *Educational Researcher*, 48(3), 158–166. <https://doi.org/10.3102/0013189X19832850>
- Maas, C. J. M., & Hox, J. J. (2004). The influence of violations of assumptions on multilevel parameter estimates and their standard errors. *Computational Statistics & Data Analysis*, 46(3), 427–440. <https://doi.org/10.1016/j.csda.2003.08.006>
- Maas, C. J. M., & Hox, J. J. (2005). Sufficient Sample Sizes for Multilevel Modeling. *Methodology*, 1(3), 86–92. <https://doi.org/10.1027/1614-1881.1.3.86>
- Maes, K., & Nijsmans, I. (1988) *Observatie in de contest von het ervaringsgericht kleuteronderwijs (EGKO): Ontwikkeling van een instrumentarium, Niet-gepubliceerde licentiaatsverhandeling*. Leuven: KU Leuven. In Laevers, F. (1994). *The Leuven Involvement Scale for Young Children, LIS-YC: Manual and video tape*. Centre for Experiential Education.
- Malaguzzi, L. (1994). *Your Image of the Child: Where Teaching Begins*.
- Malmberg, L.-E., & Martin, A. J. (2019). Processes of students' effort exertion, competence beliefs and motivation: Cyclic and dynamic effects of learning experiences within school days and school subjects. *Contemporary Educational Psychology*, 58, 299–309. <https://doi.org/10.1016/j.cedpsych.2019.03.013>

- Malmberg, L.-E., Walls, T. A., Martin, A. J., Little, T. D., & Lim, W. H. T. (2013). Primary school students' learning experiences of, and self-beliefs about competence, effort, and difficulty: Random effects models. *Learning and Individual Differences, 28*, 54–65. <https://doi.org/10.1016/j.lindif.2013.09.007>
- Mansell, W. (2007). *Education by Numbers: The tyranny of testing*. Politicos.
- Manyukhina, Y., & Wyse, D. (2019). Learner agency and the curriculum: A critical realist perspective. *The Curriculum Journal, 30*(3), 223–243. <https://doi.org/10.1080/09585176.2019.1599973>
- Mardell, B., Solis, S. L., & Bray, O. (2019). The state of play in school: Defining and promoting playful learning in formal education settings. *International Journal of Play, 8*(3), 232–236. <https://doi.org/10.1080/21594937.2019.1684157>
- Martin, A. J., Papworth, B., Ginns, P., Malmberg, L.-E., Collie, R. J., & Calvo, R. A. (2015). Real-time motivation and engagement during a month at school: Every moment of every day for every student matters. *Learning and Individual Differences, 38*, 26–35. <https://doi.org/10.1016/j.lindif.2015.01.014>
- Marulis, L. M., Baker, S. T., & Whitebread, D. (2020). Integrating metacognition and executive function to enhance young children's perception of and agency in their learning. *Early Childhood Research Quarterly, 50*, 46–54. <https://doi.org/10.1016/j.ecresq.2018.12.017>
- Mashford-Scott, A., Church, A., & Tayler, C. (2012). Seeking Children's Perspectives on their Wellbeing in Early Childhood Settings. *International Journal of Early Childhood, 44*(3), 231–247. <https://doi.org/10.1007/s13158-012-0069-7>
- Maxwell, J. A. (2012a). *A realist approach for qualitative research*. SAGE.

- Maxwell, J. A. (2012b). The Importance of Qualitative Research for Causal Explanation in Education. *Qualitative Inquiry*, 18(8), 655–661.
<https://doi.org/10.1177/1077800412452856>
- McInnes, K. (2019). Playful learning in the early years – through the eyes of children. *Education 3-13*, 47(7), 796–805. <https://doi.org/10.1080/03004279.2019.1622495>
- McIntyre, D. (2005). Bridging the gap between research and practice. *Cambridge Journal of Education*, 35(3), 357–382. <https://doi.org/10.1080/03057640500319065>
- McNally, S. A., & Slutsky, R. (2017). Key elements of the Reggio Emilia approach and how they are interconnected to create the highly regarded system of early childhood education. *Early Child Development and Care*, 187(12), 1925–1937.
<https://doi.org/10.1080/03004430.2016.1197920>
- McNeish, D. (2017). Small Sample Methods for Multilevel Modeling: A Colloquial Elucidation of REML and the Kenward-Roger Correction. *Multivariate Behavioral Research*, 52(5), 661–670. <https://doi.org/10.1080/00273171.2017.1344538>
- Mehta, J. (2015). *The Allure of Order*. Oxford University Press.
- Mercer, N., Dawes, L., & Staarman, J. K. (2009). Dialogic teaching in the primary science classroom. *Language and Education*, 23(4), 353–369.
<https://doi.org/10.1080/09500780902954273>
- Mertens, D. M. (2015). *Research and Evaluation in Education and Psychology: Integrating Diversity with Quantitative, Qualitative, and Mixed Methods* (4th Edition). Sage.
- Mick, C. (2011). Learner Agency. *European Educational Research Journal*, 10(4), 559–571.
<https://doi.org/10.2304/eeerj.2011.10.4.559>
- Midgley, C., Kaplan, A., & Middleton, M. (2001). Performance-approach goals: Good for what, for whom, under what circumstances, and at what cost? *Journal of Educational Psychology*, 93(1), 77–86. <https://doi.org/10.1037/0022-0663.93.1.77>

- Mitchell, I. (1999). Chapter 4: Bridging the Gulf Between Research and Practice. In J. Loughran (Ed.), *Researching Teaching: Methodologies and Practices for Understanding Pedagogy*. Falmer Press.
- Mitchell, M. (1993). Situational Interest: Its Multifaceted Structure in the Secondary School Mathematics Classroom. *Journal of Educational Psychology*, 85(3), 424–436.
- Molla, T., & Nolan, A. (2020). Teacher agency and professional practice. *Teachers and Teaching*, 26(1), 67–87. <https://doi.org/10.1080/13540602.2020.1740196>
- Moller, A. C., Deci, E. L., & Ryan, R. M. (2006). Choice and Ego-Depletion: The Moderating Role of Autonomy. *Personality and Social Psychology Bulletin*, 32(8), 1024–1036. <https://doi.org/10.1177/0146167206288008>
- Muller, J., & Young, M. (2019). Knowledge, power and powerful knowledge re-visited. *The Curriculum Journal*, 30(2), 196–214. <https://doi.org/10.1080/09585176.2019.1570292>
- Murray, J. (2012). Young children’s explorations: Young children’s research? *Early Child Development and Care*, 182(9), 1209–1225. <https://doi.org/10.1080/03004430.2011.604728>
- Myles, A. (2015). *Casting Light on Childism: Resisting and Transforming Prejudice and Oppression Against Children*. Widener University.
- NAHT. (2020). *Ofsted—A change for the better?* NAHT - The School Leaders’ Union. <https://naht.org.uk/News/Latest-comments/News/ArtMID/556/ArticleID/924>
- Newton, P. E., & Shaw, S. D. (2016). Disagreement over the best way to use the word ‘validity’ and options for reaching consensus. *Assessment in Education: Principles, Policy & Practice*, 23(2), 178–197. <https://doi.org/10.1080/0969594X.2015.1037241>
- Nicholl, B., & McLellan, R. (2008). ‘We’re all in this game whether we like it or not to get a number of As to Cs., Design and technology teachers’ struggles to implement

- creativity and performativity policies. *British Educational Research Journal*, 34(5), 585–600. <https://doi.org/10.1080/01411920802223875>
- Nix, G. A., Ryan, R. M., Manly, J. B., & Deci, E. L. (1999). Revitalization through Self-Regulation: The Effects of Autonomous and Controlled Motivation on Happiness and Vitality. *Journal of Experimental Social Psychology*, 35(3), 266–284. <https://doi.org/10.1006/jesp.1999.1382>
- Nurmi, J. E., & Aunola, K. (2005). Task-Motivation during the First School Years: A Person-Oriented Approach to Longitudinal Data. *Learning and Instruction*, 15(2), 103–122. <https://doi.org/10.1016/j.learninstruc.2005.04.009>
- NVivo *Qualitative Data Analysis Software* (Version 12). (1999). [Computer software]. QSR International. <https://qsrinternational.com/nvivo/nvivo-products/>
- Ofsted. (2017). *Bold beginnings: The Reception curriculum in a sample of good and outstanding primary schools* (No. 170045). Office for Standards in Education. <https://www.gov.uk/government/publications/reception-curriculum-in-good-and-outstanding-primary-schools-bold-beginnings>
- Ofsted. (2019a). *Education inspection framework: Overview of research* (No. 180045). Office for Standards in Education.
- Ofsted. (2019b). *Education inspection framework (EIF)*. Office for Standards in Education.
- O’Keefe, P. A., & Linnenbrink-Garcia, L. (2014). The role of interest in optimizing performance and self-regulation. *Journal of Experimental Social Psychology*, 53, 70–78. <https://doi.org/10.1016/j.jesp.2014.02.004>
- Oliver, M., Avramides, K., Clark, W., Hunter, J., Luckin, R., Hansen, C., & Wasson, B. (2018). Sharing teacher knowledge at scale: Teacher inquiry, learning design and the representation of teachers’ practice. *Teacher Development*, 22(4), 587–606. <https://doi.org/10.1080/13664530.2017.1381642>

- Opfer, V. D., & Pedder, D. (2011). Conceptualizing Teacher Professional Learning. *Review of Educational Research*, 81(3), 376–407. <https://doi.org/10.3102/0034654311413609>
- Osana, H. P., Lacroix, G. L., Tucker, B. J., & Desrosiers, C. (2006). The Role of Content Knowledge and Problem Features on Preservice Teachers' Appraisal of Elementary Mathematics Tasks. *Journal of Mathematics Teacher Education*, 9(4), 347–380. <https://doi.org/10.1007/s10857-006-4084-1>
- Otis, N., Grouzet, F. M. E., & Pelletier, L. G. (2005). Latent Motivational Change in an Academic Setting: A 3-Year Longitudinal Study. *Journal of Educational Psychology*, 97(2), 170–183. <https://doi.org/10.1037/0022-0663.97.2.170>
- Pakarinen, E., Lerkkanen, M.-K., & Suchodoletz, A. von. (2020). Teacher emotional support in relation to social competence in preschool classrooms. *International Journal of Research & Method in Education*, 43(4), 444–460. <https://doi.org/10.1080/1743727X.2020.1791815>
- Palincsar, A. S., Magnusson, S. J., Marano, N., Ford, D., & Brown, N. (1998). Designing a Community of Practice: Principles and Practices of the GIsML Community. *Teaching and Teacher Education*, 14(1), 5–19.
- Parker, R., & Thomsen, B. S. (2019). *Learning through play at school: A study of playful integrated pedagogies that foster children's holistic skills development in the primary school classroom*. The Lego Foundation.
- Pascal, C., & Bertram, T. (Eds.). (1997). *Effective Early Learning: Case Studies of Improvement*. Hodder and Stoughton.
- Pascal, C., & Bertram, T. (2009). Listening to young citizens: The struggle to make real a participatory paradigm in research with young children. *European Early Childhood Education Research Journal*, 17(2), 249–262. <https://doi.org/10.1080/13502930902951486>

- Pascal, C., & Bertram, T. (2012). Praxis, ethics and power: Developing praxeology as a participatory paradigm for early childhood research. *European Early Childhood Education Research Journal*, 20(4), 477–492.
<https://doi.org/10.1080/1350293X.2012.737236>
- Pascal, C., Bertram, T., Mould, C., & Hall, R. (1998). Exploring the relationship between process and outcome in young children’s learning: Stage one of a longitudinal study. *International Journal of Educational Research*, 29(1), 51–67.
[https://doi.org/10.1016/S0278-4343\(98\)00013-5](https://doi.org/10.1016/S0278-4343(98)00013-5)
- Pascal, C., Bertram, T., & Rouse, L. (2019). *Getting it right in the Early Years Foundation Stage: A review of the evidence* (p. 88). The British Association for Early Childhood Education.
- Patall, E. A., Cooper, H., & Robinson, J. C. (2008). The effects of choice on intrinsic motivation and related outcomes: A meta-analysis of research findings. *Psychological Bulletin*, 134(2), 270–300. <https://doi.org/10.1037/0033-2909.134.2.270>
- Patall, E. A., Hooper, S., Vasquez, A. C., Pituch, K. A., & Steingut, R. R. (2018). Science class is too hard: Perceived difficulty, disengagement, and the role of teacher autonomy support from a daily diary perspective. *Learning and Instruction*, 58, 220–231. <https://doi.org/10.1016/j.learninstruc.2018.07.004>
- Patall, E. A., Vasquez, A. C., Steingut, R. R., Trimble, S. S., & Pituch, K. A. (2016). Daily interest, engagement, and autonomy support in the high school science classroom. *Contemporary Educational Psychology*, 46, 180–194.
<https://doi.org/10.1016/j.cedpsych.2016.06.002>
- Pawson, R., & Tilley, N. (1997). *Realistic Evaluation*. SAGE Publications Ltd.
- Pekrun, R., Goetz, T., Daniels, L. M., Stupnisky, R. H., & Perry, R. P. (2010). Boredom in achievement settings: Exploring control–value antecedents and performance

- outcomes of a neglected emotion. *Journal of Educational Psychology*, 102(3), 531–549. <https://doi.org/10.1037/a0019243>
- Peressini, D., Borko, H., Romagnano, L., Knuth, E., & Willis, C. (2004). A Conceptual Framework for Learning to Teach Secondary Mathematics: A Situative Perspective. *Educational Studies in Mathematics*, 56(1), 67–96. <https://doi.org/10.1023/B:EDUC.0000028398.80108.87>
- Perry, N. E. (1998). Young children's self-regulated learning and contexts that support it. *Journal of Educational Psychology*, 90(4), 715–729. <https://doi.org/10.1037/0022-0663.90.4.715>
- Perry, N. E. (2013). Understanding classroom processes that support children's self-regulation of learning. *Self-Regulation and Dialogue in Primary Classrooms*, 11(10), 45–68.
- Perry, N. E., & VandeKamp, K. J. O. (2000). Creating classroom contexts that support young children's development of self-regulated learning. *International Journal of Educational Research*, 33(7–8), 821–843. [https://doi.org/10.1016/S0883-0355\(00\)00052-5](https://doi.org/10.1016/S0883-0355(00)00052-5)
- Perry, N. E., Walton, C., & Calder, K. (1999). Teachers Developing Assessments of Early Literacy: A Community of Practice Project. *Teacher Education and Special Education: The Journal of the Teacher Education Division of the Council for Exceptional Children*, 22(4), 218–233. <https://doi.org/10.1177/088840649902200404>
- Ponitz, C. C., Rimm-Kaufman, S. E., Grimm, K. J., & Curby, T. W. (2009). Kindergarten classroom quality, behavioral engagement, and reading achievement. *School Psychology Review*, 38(1), 102.
- Portilla, X. A., Ballard, P. J., Adler, N. E., Boyce, W. T., & Obradović, J. (2014). An Integrative View of School Functioning: Transactions between Self-Regulation,

- School Engagement, and Teacher-Child Relationship Quality. *Child Development*, 85(5), 1915–1931. <https://doi.org/10.1111/cdev.12259>
- Powell, D. R., Burchinal, M. R., File, N., & Kontos, S. (2008). An eco-behavioral analysis of children's engagement in urban public school preschool classrooms. *Early Childhood Research Quarterly*, 23(1), 108–123. <https://doi.org/10.1016/j.ecresq.2007.04.001>
- Powell, M. A., & Smith, A. B. (2009). Children's Participation Rights in Research. *Childhood*, 16(1), 124–142. <https://doi.org/10.1177/0907568208101694>
- Power, S., Rhys, M., Taylor, C., & Waldron, S. (2019). How child-centred education favours some learners more than others. *Review of Education*, 7(3), 570–592. <https://doi.org/10.1002/rev3.3137>
- Purdon, A. (2013). How practitioners promote sustained shared thinking. *Early Years Educator*, 15(8), 38–44. <https://doi.org/10.12968/eyed.2013.15.8.38>
- Putnam, S. P., & Rothbart, M. K. (2006). Development of Short and Very Short Forms of the Children's Behavior Questionnaire. *Journal of Personality Assessment*, 87(1), 102–112. https://doi.org/10.1207/s15327752jpa8701_09
- Pyle, A., & Danniels, E. (2017). A Continuum of Play-Based Learning: The Role of the Teacher in Play-Based Pedagogy and the Fear of Hijacking Play. *Early Education and Development*, 28(3), 274–289. <https://doi.org/10.1080/10409289.2016.1220771>
- R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>
- Rainio, A. P., & Hilppö, J. (2017). The dialectics of agency in educational ethnography. *Ethnography and Education*, 12(1), 78–94. <https://doi.org/10.1080/17457823.2016.1159971>

- Reeve, J. (2009). Why Teachers Adopt a Controlling Motivating Style Toward Students and How They Can Become More Autonomy Supportive. *Educational Psychologist*, 44(3), 159–175. <https://doi.org/10.1080/00461520903028990>
- Reeve, J., & Cheon, S. H. (2016). Teachers become more autonomy supportive after they believe it is easy to do. *Psychology of Sport and Exercise*, 22, 178–189. <https://doi.org/10.1016/j.psychsport.2015.08.001>
- Reeve, J., & Jang, H. (2006). What teachers say and do to support students' autonomy during a learning activity. *Journal of Educational Psychology*, 98(1), 209–218. <https://doi.org/10.1037/0022-0663.98.1.209>
- Reeve, J., Jang, H., Carrell, D., Jeon, S., & Barch, J. (2004). Enhancing students' engagement by increasing teachers' autonomy support. *Motivation and Emotion*, 28(2), 147–169.
- Reeve, J., Nix, G., & Hamm, D. (2003). Testing models of the experience of self-determination in intrinsic motivation and the conundrum of choice. *Journal of Educational Psychology*, 95(2), 375–392. <https://doi.org/10.1037/0022-0663.95.2.375>
- Reeve, J., & Tseng, C.-M. (2011). Agency as a fourth aspect of students' engagement during learning activities. *Contemporary Educational Psychology*, 36(4), 257–267. <https://doi.org/10.1016/j.cedpsych.2011.05.002>
- Reggio Children: International Center for the Defense and Promotion of the Rights and Potential of all Children & Harvard Project Zero. (2001). *Making learning visible: Children as individual and group learners*. Reggio Emilia: Reggio Children.
- Renaud-Dubé, A., Guay, F., Talbot, D., Taylor, G., & Koestner, R. (2015). The relations between implicit intelligence beliefs, autonomous academic motivation, and school persistence intentions: A mediation model. *Social Psychology of Education*, 18(2), 255–272. <https://doi.org/10.1007/s11218-014-9288-0>

- Renninger, K. A., & Bachrach, J. E. (2015). Studying Triggers for Interest and Engagement Using Observational Methods. *Educational Psychologist, 50*(1), 58–69.
<https://doi.org/10.1080/00461520.2014.999920>
- Renninger, K. A., Bachrach, J. E., & Hidi, S. E. (2018). Triggering and maintaining interest in early phases of interest development. *Learning, Culture and Social Interaction*.
<https://doi.org/10.1016/j.lcsi.2018.11.007>
- Renninger, K. A., & Hidi, S. (2011). Revisiting the Conceptualization, Measurement, and Generation of Interest. *Educational Psychologist, 46*(3), 168–184.
<https://doi.org/10.1080/00461520.2011.587723>
- Renninger, K. A., Hidi, S., & Krapp, A. (1992). *The Role of Interest in Learning and Development*. Psychology Press.
- Rimm-Kaufman, S. E., Early, D. M., Cox, M. J., Saluja, G., Pianta, R. C., Bradley, R. H., & Payne, C. (2002). Early behavioral attributes and teachers' sensitivity as predictors of competent behavior in the kindergarten classroom. *Journal of Applied Developmental Psychology, 23*(4), 451–470. [https://doi.org/10.1016/S0193-3973\(02\)00128-4](https://doi.org/10.1016/S0193-3973(02)00128-4)
- Rinaldi, C. (2001). The Pedagogy of Listening: The Listening Perspective from Reggio Emilia. *Innovations in Early Education: The International Reggio Exchange, 8*(4), 1–4.
- Roberts, J. (2021). Teachers' leader slams Ofsted for 'reign of terror'. *Tes Magazine*. Retrieved 17 January 2022, from <https://www.tes.com/magazine/news/general/teachers-leader-slams-ofsted-reign-terror>
- Roberts, J. K., Monaco, J. P., Stovall, H., & Foster, V. (2010). *Explained Variance in Multilevel Models*. Routledge Handbooks Online.
<https://doi.org/10.4324/9780203848852.ch12>

- Robson, S. (2016). Self-regulation, metacognition and child- and adult-initiated activity: Does it matter who initiates the task? *Early Child Development and Care*, 186(5), 764–784. <https://doi.org/10.1080/03004430.2015.1057581>
- Roche, E. (1996). The Bumpy Bus Ride to the Democratic Classroom. In L. E. Beyer (Ed.), *Creating Democratic Classrooms* (pp. 27–40). Teacher College Press.
- Roebbers, C. M. (2017). Executive function and metacognition: Towards a unifying framework of cognitive self-regulation. *Developmental Review*, 45, 31–51. <https://doi.org/10.1016/j.dr.2017.04.001>
- Rojas-Drummond, S., Torreblanca, O., Pedraza, H., Vélez, M., & Guzmán, K. (2013). ‘Dialogic scaffolding’: Enhancing learning and understanding in collaborative contexts. *Learning, Culture and Social Interaction*, 2(1), 11–21. <https://doi.org/10.1016/j.lcsi.2012.12.003>
- Rotgans, J. I., & Schmidt, H. G. (2011). Situational interest and academic achievement in the active-learning classroom. *Learning and Instruction*, 21(1), 58–67. <https://doi.org/10.1016/j.learninstruc.2009.11.001>
- Rothbaum, F., & Trommsdorff, G. (2007). Do roots and wings complement or oppose one another?: The socialization of relatedness and autonomy in cultural context. In *Handbook of socialization: Theory and research*. (pp. 461–489). The Guilford Press. <https://psycnet.apa.org/record/2006-23344-018>
- Ruiz-Gallardo, J.-R., Verde, A., & Valdés, A. (2013). Garden-Based Learning: An Experience With “At Risk” Secondary Education Students. *The Journal of Environmental Education*, 44(4), 252–270. <https://doi.org/10.1080/00958964.2013.786669>
- Ryan, R. M. (2017). *Self-determination theory basic psychological needs in motivation, development, and wellness*. New York: Guilford Press.

- Ryan, R. M., & Connell, J. P. (1989). Perceived locus of causality and internalization: Examining reasons for acting in two domains. *Journal of Personality and Social Psychology*, 57(5), 749–761. <https://doi.org/10.1037//0022-3514.57.5.749>
- Ryan, R. M., Connell, J. P., & Plant, R. W. (1990). Emotions in nondirected text learning. *Learning and Individual Differences*, 2(1), 1–17. [https://doi.org/10.1016/1041-6080\(90\)90014-8](https://doi.org/10.1016/1041-6080(90)90014-8)
- Ryan, R. M., & Deci, E. L. (2000). Intrinsic and Extrinsic Motivations: Classic Definitions and New Directions. *Contemporary Educational Psychology*, 25(1), 54–67. <https://doi.org/10.1006/ceps.1999.1020>
- Ryan, R. M., Deci, E. L., Grolnick, W. S., & La Guardia, J. G. (2006). The significance of autonomy and autonomy support in psychological development and psychopathology. In *Developmental psychopathology: Theory and method, Vol. 1, 2nd ed* (pp. 795–849). John Wiley & Sons, Inc.
- Ryan, R. M., Huta, V., & Deci, E. L. (2008). Living well: A self-determination theory perspective on eudaimonia. *Journal of Happiness Studies: An Interdisciplinary Forum on Subjective Well-Being*, 9(1), 139–170. <https://doi.org/10.1007/s10902-006-9023-4>
- Ryan, R. M., & Lynch, M. (2003). Motivation and Classroom Management. In R. Curren (Ed.), *A companion to the philosophy of education* (pp. 260–271). Blackwell. <https://ebookcentral.proquest.com/lib/cam/reader.action?docID=214150#>
- Rycroft-Smith, L., & Dutaut, J.-L. (2017). *Flip the system UK : a teachers' manifesto*. London : Routledge, 2017.
- Sahlberg, P. (2016). The Global Educational Reform Movement and Its Impact on Schooling. In Karen Mundy, Andy Green, Bob Lingard, & Antoni Verger (Eds.), *The handbook of global education policy* (pp. 128–144). John Wiley & Sons, Inc.

<https://ebookcentral.proquest.com/lib/CAM/detail.action?docID=4427184&pq-origsite=primo>

- Sak, R., Erden, F. T., & Morrison, G. S. (2016). Child-centred education: Preschool teachers' beliefs and self-reported practices. *Early Child Development and Care*, 186(8), 1185–1202. <https://doi.org/10.1080/03004430.2015.1081185>
- Salkind, N. (Ed.). (2010). Krippendorff's Alpha. In *Encyclopedia of Research Design*. SAGE Publications, Inc. <https://doi.org/10.4135/9781412961288.n206>
- Salmela-Aro, K., Moeller, J., Schneider, B., Spicer, J., & Lavonen, J. (2016). Integrating the light and dark sides of student engagement using person-oriented and situation-specific approaches. *Learning and Instruction*, 43, 61–70. <https://doi.org/10.1016/j.learninstruc.2016.01.001>
- Samuelsson, I. P., & Carlsson, M. A. (2008). The Playing Learning Child: Towards a pedagogy of early childhood. *Scandinavian Journal of Educational Research*, 52(6), 623–641. <https://doi.org/10.1080/00313830802497265>
- Sansone, C., & Thoman, D. B. (2005). Interest as the Missing Motivator in Self-Regulation. *European Psychologist*, 10(3), 175–186. <https://doi.org/10.1027/1016-9040.10.3.175>
- Sawyer, J. (2017). I think I can: Preschoolers' private speech and motivation in playful versus non-playful contexts. *Early Childhood Research Quarterly*, 38, 84–96. <https://doi.org/10.1016/j.ecresq.2016.09.004>
- Sayer, A. (1997). Critical Realism and the Limits to Critical Social Science. *Journal for the Theory of Social Behaviour*, 27(4), 473–488. <https://doi.org/10.1111/1468-5914.00052>
- Schiefele, U., & Krapp, A. (1996). Topic interest and free recall of expository text. *Learning and Individual Differences*, 8(2), 141–160. [https://doi.org/10.1016/S1041-6080\(96\)90030-8](https://doi.org/10.1016/S1041-6080(96)90030-8)

- Schindler, H., & Berry, M. (2017, October 31). *Introduction to the IDEAS impact framework*. Frontiers of Innovation, Center on the Developing Child at Harvard University, Cambridge MA.
- Schukajlow, S., & Rakoczy, K. (2016). The power of emotions: Can enjoyment and boredom explain the impact of individual preconditions and teaching methods on interest and performance in mathematics? *Learning and Instruction, 44*, 117–127.
<https://doi.org/10.1016/j.learninstruc.2016.05.001>
- Schurz, G. (2008). Patterns of abduction. *Synthese, 164*(2), 201–234.
<https://doi.org/10.1007/s11229-007-9223-4>
- Schwartz, B. (2000). Self-determination: The tyranny of freedom. *American Psychologist, 55*(1), 79–88. <https://doi.org/10.1037//0003-066X.55.1.79>
- Seigel, S. (2001). To Be Part of Something Bigger than Oneself. In Reggio Children: International Center for the Defense and Promotion of the Rights and Potential of all Children & Harvard Project Zero, *Making learning visible: Children as individual and group learners*. Reggio Emili: Reggio Children.
- Sherman, B., & Teemant, A. (2021). Unravelling effective professional development: A rhizomatic inquiry into coaching and the active ingredients of teacher learning. *Professional Development in Education, 47*(2–3), 363–376.
<https://doi.org/10.1080/19415257.2020.1825511>
- Shernoff, D. J., Csikszentmihalyi, M., Schneider, B., & Shernoff, E. S. (2003). Student Engagement in High School Classrooms from the Perspective of Flow Theory. *School Psychology Quarterly, 18*(2), 158–176.
- Siddiqui, N., Gorard, S., & See, B. H. (2018). The importance of process evaluation for randomised control trials in education. *Educational Research, 60*(3), 357–370.
<https://doi.org/10.1080/00131881.2018.1493349>

- Sierens, E., Vansteenkiste, M., Goossens, L., Soenens, B., & Dochy, F. (2009). The synergistic relationship of perceived autonomy support and structure in the prediction of self-regulated learning. *British Journal of Educational Psychology*, 79(1), 57–68. <https://doi.org/10.1348/000709908X304398>
- Simpson, A. (2017). The misdirection of public policy: Comparing and combining standardised effect sizes. *Journal of Education Policy*, 32(4), 450–466. <https://doi.org/10.1080/02680939.2017.1280183>
- Sims, S., Fletcher-Wood, H., O'Mara, A., Cottingham, S., Stansfield, C., Herwegen, J. V., & Anders, J. (2021). *What are the Characteristics of Effective Teacher Professional Development? A Systematic Review and Meta-analysis* (p. 196). Education Endowment Foundation.
- Singer, D. G., Golinkoff, R. M., & Hirsh-Pasek, K. (2006). *Play = learning how play motivates and enhances children's cognitive and social-emotional*.
- Siraj-Blatchford, I. (2010). *A focus on pedagogy: Case studies of effective practice*. In K. Sylva, E. Melhuish, P. Sammons, I. Siraj-Blatchford, & B. Taggart (Eds.), *Early Childhood Matters: Evidence from the Effective Pre-school and Primary Education project* (pp. 149–165). Routledge.
- Siraj, I., & Asani, R. (2014). The role of sustained shared thinking, play and metacognition in young children's learning. In *The Routledge International Handbook of Young Children's Thinking and Understanding*. Routledge.
- Siraj-Blatchford, I., & Sylva, K. (2004). Researching pedagogy in English pre-schools. *British Educational Research Journal*, 30(5), 713–730. <https://doi.org/10.1080/0141192042000234665>
- Siraj-Blatchford, I., Taggart, B., Sylva, K., Sammons, P., & Melhuish, E. (2008). Towards the transformation of practice in early childhood education: The effective provision of

- pre-school education (EPPE) project. *Cambridge Journal of Education*, 38(1), 23–36.
<https://doi.org/10.1080/03057640801889956>
- Skene, K., O’Farrelly, C. M., Byrne, E. M., Kirby, N., Stevens, E. C., & Ramchandani, P. G. (2022). Can guidance during play enhance children’s learning and development in educational contexts? A systematic review and meta-analysis. *Child Development*, 1–19. <https://doi.org/10.1111/cdev.13730>
- Skinner, E. A., & Belmont, M. J. (1993). Motivation in the classroom: Reciprocal effects of teacher behavior and student engagement across the school year. *Journal of Educational Psychology*, 85(4), 571–581. <https://doi.org/10.1037/0022-0663.85.4.571>
- Snijders, T. A. B. (2005). Power and Sample Size in Multilevel Linear Models. In *Encyclopedia of Statistics in Behavioral Science*. American Cancer Society.
<https://doi.org/10.1002/0470013192.bsa492>
- Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel analysis: An introduction to basic and advanced multilevel modeling / Tom A.B. Snijders and Roel J. Bosker*. (2nd ed.). Los Angeles.
- Spinath, B., & Spinath, F. M. (2005). Longitudinal analysis of the link between learning motivation and competence beliefs among elementary school children. *Learning and Instruction*, 15(2), 87–102. <https://doi.org/10.1016/j.learninstruc.2005.04.008>
- Stahl, A. E., & Feigenson, L. (2015). Observing the unexpected enhances infants’ learning and exploration. *Science*. <https://doi.org/10.1126/science.aaa3799>
- StataCorp. (2019). *Stata Statistical Software: Release 16*. College Station, TX: StataCorp LLC.
- Stefanou, C. R., Perencevich, K. C., DiCintio, M., & Turner, J. C. (2004). Supporting autonomy in the classroom: Ways teachers encourage student decision making and

ownership. *Educational Psychologist*, 39(2), 97–110.

https://doi.org/10.1207/s15326985ep3902_2

- Stefansson, K. K., Gestsdottir, S., Birgisdottir, F., & Lerner, R. M. (2018). School engagement and intentional self-regulation: A reciprocal relation in adolescence. *Journal of Adolescence*, 64, 23–33. <https://doi.org/10.1016/j.adolescence.2018.01.005>
- Stipek, D., Feiler, R., Daniels, D., & Milburn, S. (1995). Effects of Different Instructional Approaches on Young Children’s Achievement and Motivation. *Child Development*, 66(1), 209. <https://doi.org/10.2307/1131201>
- Stoppard, T. (2013). *Shipwreck Tom Stoppard*. London : Bloomsbury, c2013.
- Stumm, S. von, Hell, B., & Chamorro-Premuzic, T. (2011). The Hungry Mind Intellectual Curiosity Is the Third Pillar of Academic Performance. *Perspectives on Psychological Science*, 6(6), 574–588. <https://doi.org/10.1177/1745691611421204>
- Sturrock, S. (2021). Primary teachers’ experiences of neo-liberal education reform in England: ‘Nothing is ever good enough’. *Research Papers in Education*, 0(0), 1–27. <https://doi.org/10.1080/02671522.2021.1941213>
- Swennen, A., Lunenberg, M., & Korthagen, F. (2008). Preach what you teach! Teacher educators and congruent teaching. *Teachers and Teaching*, 14(5–6), 531–542. <https://doi.org/10.1080/13540600802571387>
- Sylva, K., Sammons, P., Melhuish, E., Siraj, I., & Taggart, B. (2020). Developing 21st century skills in early childhood: The contribution of process quality to self-regulation and pro-social behaviour. *Zeitschrift Für Erziehungswissenschaft*. <https://doi.org/10.1007/s11618-020-00945-x>
- Taber, K. (2013). *Classroom based Research and Evidence-based Practice: An introduction* (2nd ed.). Sage.

- TACTYC. (2017). *Bald Beginnings. A Response to Ofsted's (2017) report, Bold beginnings: The Reception curriculum in a sample of good and outstanding primary schools.* (Association for Professional Development in Early Years. <http://tactyc.org.uk/wp-content/uploads/2017/12/Bold-Beginnings-TACTYC-response-FINAL-09.12.17.pdf>)
- Taffoni, F., Tamilia, E., Focaroli, V., Formica, D., Ricci, L., Pino, G. D., Baldassarre, G., Mirolli, M., Guglielmelli, E., & Keller, F. (2014). Development of goal-directed action selection guided by intrinsic motivations: An experiment with children. *Experimental Brain Research*, 232(7), 2167–2177. <https://doi.org/10.1007/s00221-014-3907-z>
- Tarnanen, M., Kostianen, E., Kaukonen, V., Martin, A., & Toikka, T. (2021). Towards a learning community: Understanding teachers' mental models to support their professional development and learning. *Professional Development in Education*, 0(0), 1–15. <https://doi.org/10.1080/19415257.2021.1959383>
- Taylor, G., Jungert, T., Mageau, G. A., Schattke, K., Dedic, H., Rosenfield, S., & Koestner, R. (2014). A self-determination theory approach to predicting school achievement over time: The unique role of intrinsic motivation. *Contemporary Educational Psychology*, 39(4), 342–358. <https://doi.org/10.1016/j.cedpsych.2014.08.002>
- Thomas, A., & Pattison, H. (2013). Informal Home Education: Philosophical Aspirations put into Practice. *Studies in Philosophy and Education*, 32(2), 141–154. <https://doi.org/10.1007/s11217-012-9299-2>
- Thomas, N. (2021). Childhood and Participation: Between domination and freedom. In R. Braches-Chyrek (Ed.), *The Future of Childhood Studies* (1st ed.). Verlag Barbara Budrich. <https://doi.org/10.2307/j.ctv1dhpgx1>

- Tillema, H. H. (2000). Belief change towards self-directed learning in student teachers: Immersion in practice or reflection on action. *Teaching and Teacher Education*, 16(5), 575–591. [https://doi.org/10.1016/S0742-051X\(00\)00016-0](https://doi.org/10.1016/S0742-051X(00)00016-0)
- Timperley, H., Kaser, L., & Halbert, J. (2014). *A framework for transforming learning in schools: Innovation and the spiral of inquiry*. Victoria, Australia: Centre for Strategic Education.
- Tizard, B., & Hughes, M. (1984). *Children's Learning at Home and in School*. Fontana.
- Torres, P. E., Whitebread, D., & McLellan, R. (2018). The Role of Teacher Regulatory Talk in Students' Self-Regulation Development Across Cultures. *New Directions for Child and Adolescent Development*, 2018(162), 89–114. <https://doi.org/10.1002/cad.20259>
- Traxler, C. R. (2015). The Most Democratic School of Them All: Why the Sudbury Model of Education Should Be Taken Seriously. *Schools: Studies in Education*, 12(2), 271–296. <https://doi.org/10.1086/683220>
- Trochim, W. M. K. (2016). *Research methods the essential knowledge base*. Cengage Learning.
- Trollope, R. (2017). *Teacher Understanding and Application of Continuous Provision in Key Stage One and Key Stage Two*. University of Cambridge.
- Tulis, M., & Fulmer, S. M. (2013). Students' motivational and emotional experiences and their relationship to persistence during academic challenge in mathematics and reading. *Learning and Individual Differences*, 27, 35–46. <https://doi.org/10.1016/j.lindif.2013.06.003>
- Tzuo, P. W. (2007). The Tension between Teacher Control and Children's Freedom in a Child-centered Classroom: Resolving the Practical Dilemma through a Closer Look at the Related Theories. *Early Childhood Education Journal*, 35(1), 33–39. <https://doi.org/10.1007/s10643-007-0166-7>

- Ulich, M., & Mayr, T. (2002). Children's involvement profiles in daycare centres. *European Early Childhood Education Research Journal*, 10(2), 127–143.
<https://doi.org/10.1080/13502930285209001>
- UNGA (United Nations General Assembly) (1989) *Convention on the Rights of the Child*.
UN General Assembly. Resolution 44/25 of November 20, 1989. Available at:
<https://www.unicef.org.uk/what-we-do/un-convention-child-rights/>
- Uprichard, E. (2008). Children as 'Being and Becomings': Children, Childhood and Temporality. *Children & Society*, 22(4), 303–313. <https://doi.org/10.1111/j.1099-0860.2007.00110.x>
- Valeski, T. N., & Stipek, D. J. (2001). Young Children's Feelings about School. *Child Development*, 72(4), 1198–1213. <https://doi.org/10.1111/1467-8624.00342>
- Vallerand, R. J., & Bissonnette, R. (1992). Intrinsic, Extrinsic, and Amotivational Styles as Predictors of Behavior: A Prospective Study. *Journal of Personality*, 60(3), 599–620.
<https://doi.org/10.1111/j.1467-6494.1992.tb00922.x>
- Vallerand, R. J., Fortier, M. S., & Guay, F. (1997). Self-Determination and Persistence in a Real-Life Setting Toward a Motivational Model of High School Dropout. *Journal of Personality and Social Psychology*, 72(5), 1161–1176.
- Valli, L., & Chambliss, M. (2007). Creating Classroom Cultures: One Teacher, Two Lessons, and a High-Stakes Test. *Anthropology & Education Quarterly*, 38(1), 57–75.
<https://doi.org/10.1525/aeq.2007.38.1.57>
- van Buuren, S. (2010). *Multiple Imputation of Multilevel Data*. Routledge Handbooks Online. <https://doi.org/10.4324/9780203848852.ch10>
- Vangrieken, K., Meredith, C., Packer, T., & Kyndt, E. (2017). Teacher communities as a context for professional development: A systematic review. *Teaching and Teacher Education*, 61, 47–59. <https://doi.org/10.1016/j.tate.2016.10.001>

- Vansteenkiste, M., Sierens, E., Goossens, L., Soenens, B., Dochy, F., Mouratidis, A., Aelterman, N., Haerens, L., & Beyers, W. (2012). Identifying configurations of perceived teacher autonomy support and structure: Associations with self-regulated learning, motivation and problem behavior. *Learning and Instruction*, 22(6), 431–439. <https://doi.org/10.1016/j.learninstruc.2012.04.002>
- Vasalampi, K., Muotka, J., Malmberg, L.-E., Aunola, K., & Lerkkanen, M.-K. (2020). Intra-individual dynamics of lesson-specific engagement: Lagged and cross-lagged effects from one lesson to the next. *British Journal of Educational Psychology*, n/a(n/a), e12404. <https://doi.org/10.1111/bjep.12404>
- Vermunt, J. D., Vrikki, M., van Halem, N., Warwick, P., & Mercer, N. (2019). The impact of Lesson Study professional development on the quality of teacher learning. *Teaching and Teacher Education*, 81, 61–73. <https://doi.org/10.1016/j.tate.2019.02.009>
- Viseu, F., & Oliveira, I. B. (2012). Open-ended Tasks in the Promotion of Classroom Communication in Mathematics. *International Electronic Journal of Elementary Education*, 4(2), 287–300 .
- Vitiello, V. E., Booren, L. M., Downer, J. T., & Williford, A. P. (2012). Variation in children's classroom engagement throughout a day in preschool: Relations to classroom and child factors. *Early Childhood Research Quarterly*, 27(2), 210–220. <https://doi.org/10.1016/j.ecresq.2011.08.005>
- Vrikki, M., Warwick, P., Vermunt, J. D., Mercer, N., & Van Halem, N. (2017). Teacher learning in the context of Lesson Study: A video-based analysis of teacher discussions. *Teaching and Teacher Education*, 61, 211–224. <https://doi.org/10.1016/j.tate.2016.10.014>

- Vrikki, M., Wheatley, L., Howe, C., Hennessy, S., & Mercer, N. (2019). Dialogic practices in primary school classrooms. *Language and Education*, 33(1), 85–100.
<https://doi.org/10.1080/09500782.2018.1509988>
- Waller, T., & Bitou, A. (2011). Research with children: Three challenges for participatory research in early childhood. *European Early Childhood Education Research Journal*, 19(1), 5–20. <https://doi.org/10.1080/1350293X.2011.548964>
- Walsh, G., Sproule, L., McGuinness, C., Trew, K., Rafferty, H., & Sheehy, N. (2006). An appropriate curriculum for 4–5-year-old children in Northern Ireland: Comparing play-based and formal approaches. *Early Years*, 26(2), 201–221.
<https://doi.org/10.1080/09575140600760003>
- Warwick, P., Mercer, N., & Kershner, R. (2013). ‘Wait, let’s just think about this’: Using the interactive whiteboard and talk rules to scaffold learning for co-regulation in collaborative science activities. *Learning, Culture and Social Interaction*, 2(1), 42–51. <https://doi.org/10.1016/j.lcsi.2012.12.004>
- Watson, S. (2020). New Right 2.0: Teacher populism on social media in England. *British Educational Research Journal*. <https://doi.org/10.1002/berj.3664>
- Webel, C., & Platt, D. (2015). The role of professional obligations in working to change one’s teaching practices. *Teaching and Teacher Education*, 47, 204–217.
<https://doi.org/10.1016/j.tate.2015.01.007>
- Weisberg, D. S., Hirsh-Pasek, K., & Golinkoff, R. M. (2013). Guided Play: Where Curricular Goals Meet a Playful Pedagogy. *Mind, Brain, and Education*, 7(2), 104–112.
<https://doi.org/10.1111/mbe.12015>
- Whitebread, D. (2018). Play: The new renaissance. *International Journal of Play*, 7(3), 237–243. <https://doi.org/10.1080/21594937.2018.1532952>

- Whitebread, D., Basilio, M., Kvalja, M., & Verma, M. (2012). The importance of play. *Brussels, Belgium: Toys Industries for Europe*.
http://www.importanceofplay.eu/IMG/pdf/dr_david_whitebread_-_the_importance_of_play.pdf
- Wilde, M., Basten, M., Großmann, N., Haunhorst, D., Desch, I., Strüber, M., & Randler, C. (2018). The (non-)benefit of choosing: If you get what you want it is not important that you chose it. *Motivation and Emotion, 42*(3), 348–359.
<https://doi.org/10.1007/s11031-018-9675-5>
- Williams, D. H. J. (2021, October 21). *Exploratory maths and dedicated maths sessions. How both are important*. Ponderings on Maths.
<https://info125328.wixsite.com/website/post/exploratory-maths-and-dedicated-maths-sessions-how-both-are-important>
- Williamson, B., & Morgan, J. (2009). Educational reform, enquiry-based learning and the re-professionalisation of teachers. *The Curriculum Journal, 20*(3), 287–304.
<https://doi.org/10.1080/09585170903195894>
- Wilson, M. A. F. (2015). Radical democratic schooling on the ground: Pedagogical ideals and realities in a Sudbury school. *Ethnography and Education, 10*(2), 121–136.
<https://doi.org/10.1080/17457823.2014.959978>
- Wood, E. (2019). Unbalanced and unbalancing acts in the Early Years Foundation Stage: A critical discourse analysis of policy-led evidence on teaching and play from the office for standards in education in England (Ofsted). *Education 3-13, 0*(0), 1–12.
<https://doi.org/10.1080/03004279.2019.1622494>
- Wood, E. A. (2014). Free choice and free play in early childhood education: Troubling the discourse. *International Journal of Early Years Education, 22*(1), 4–18.
<https://doi.org/10.1080/09669760.2013.830562>

- Wood, E., & Hedges, H. (2016). Curriculum in early childhood education: Critical questions about content, coherence, and control. *The Curriculum Journal*, 27(3), 387–405.
<https://doi.org/10.1080/09585176.2015.1129981>
- Wrigley, T. (2018). The power of ‘evidence’: Reliable science or a set of blunt tools? *British Educational Research Journal*, 44(3), 359–376. <https://doi.org/10.1002/berj.3338>
- Wyse, D., & Torgerson, C. (2017). Experimental trials and ‘what works?’ in education: The case of grammar for writing. *British Educational Research Journal*, 43(6), 1019–1047. <https://doi.org/10.1002/berj.3315>
- Yanow, D. (2015). Neither rigorous nor objective? Interrogating criteria for knowledge claims in interpretive science. In D. Yanow & P. Schwartz-Shea (Eds.), *Interpretation and method: Empirical research methods and the interpretive turn* (2nd ed., pp. 97–119). Routledge.
<https://ebookcentral.proquest.com/lib/cam/detail.action?docID=302458>
- Yazzie-Mintz, E. (2010). *Charting the Path from Engagement to Achievement: A Report on the 2009 High School Survey of Student Engagement*. Indiana University Center for Evaluation & Education Policy. <http://hub.mspnet.org/index.cfm/20806>
- Young, R. (1986). *Personal autonomy: Beyond negative and positive liberty*. Croom Helm.
- Zahorik, J. A. (1996). Elementary and Secondary Teachers’ Reports of How They Make Learning Interesting. *The Elementary School Journal*, 96(5), 551–564.
- Zapf, A., Castell, S., Morawietz, L., & Karch, A. (2016). Measuring inter-rater reliability for nominal data – which coefficients and confidence intervals are appropriate? *BMC Medical Research Methodology*, 16(1), 93. [https://doi.org/10.1186/s12874-016-0200-](https://doi.org/10.1186/s12874-016-0200-9)

- Zosh, J. M., Hirsh-Pasek, K., Hopkins, E. J., Jensen, H., Liu, C., Neale, D., Solis, S. L., & Whitebread, D. (2018). Accessing the Inaccessible: Redefining Play as a Spectrum. *Frontiers in Psychology, 9*. <https://doi.org/10.3389/fpsyg.2018.01124>
- Zyngier, D. (2008). (Re)conceptualising student engagement: Doing education not doing time. *Teaching and Teacher Education, 24*(7), 1765–1776. <https://doi.org/10.1016/j.tate.2007.09.004>

Appendices

Appendix A - Stepping Stones teacher-researcher Memorandum of Understanding

25th March 2019

Dear _____ ,

Thank you very much for agreeing to participate in our research project “Stepping Stones in Science”. We really appreciate your willingness to support our work in your school. The project will bring together teachers from multiple schools, establishing a community of practice to explore and implement research on supporting children’s autonomy in their learning. We hope you will benefit from this experience, which will help us learn how we can help teachers foster children’s motivation and self-regulation.

The project will be in the form of a series of workshops between one day and a few hours in length. We are flexible and we want to make sure that your participation in the project is rewarding for you and the children. The different components of the project are:

- **Community of Practice sessions:** There will be one whole day workshop to introduce the project and the research. This will be followed by three or four afternoon workshops. These will most likely take place at the Faculty of Education or one of the participating schools. Costs to do with travel and teaching cover will be reimbursed. Some parts of the community of practice sessions will be audio-recorded. Teachers may also be invited to answer questions on video if they wish to. Researchers will make copies of works produced to help inform our research.
- **Discussions with teachers:** A research assistant will visit your class at the start, middle and end of the project to chat with you. This is to help us understand the context in which you work, as well as your thoughts on what works well when translating the research into real classrooms. This conversation will be audio-recorded to help inform our research.
- **Class visits:** A research assistant will visit your class at the start, middle and end of the project to see which aspects of the ideas and materials teachers are using most. These visits will focus on one science lesson and one literacy or numeracy lesson wherever possible. The researcher will simply watch quietly from a spot that is convenient in the classroom and make some notes. Please note that this is not an assessment of your performance as a teacher and we do not ask that you teach anything other than what you had planned for that day. The aim is to better understand teachers and their practice. At the end of the day, teachers will be invited to share their aims and thoughts about the lesson with the researcher so that we understand your perspective.

Our project has been approved by the Faculty of Education and follows GDPR data protection guidelines. All information you share with us will remain confidential. We replace names with codes or pseudonyms in our files, so research information can never be linked directly to any individual's identity. Anonymised data (e.g. names replaced with codes) may be shared with other researchers and made open for re-use. If at any point you have concerns that comments make you or another person in your school recognisable, please share these concerns with us and we will jointly consider steps to address them.

We aim to put teachers in the driving seat of their own professional development. You can choose which practices you would like to try and how to implement them, and are under no obligation to make changes if that is your preference. Your thoughts on the project are very important to us, and so you may also be invited to comment on our reports to ensure you are happy with what we choose to share.

We hope that you will find the project an enriching experience. In the past, teachers who have joined us as co-researchers have found the experience highly rewarding, because it allows them time to reflect on their own practice, to share practices with other teachers, and to explore how state-of-the-art research evidence can be adapted to work in their own context, all as part of a supportive community. From our point of view, the project is a chance to continue building an evidence-informed programme of professional development with teachers and for teachers, that can lead to positive changes in children's autonomy.

We are always available to answer any questions you may have and to discuss further details so please do not hesitate to contact us. We truly appreciate your support, and believe that together we can contribute to our understanding of children's development and learning.

Yours sincerely,

The Stepping Stones team

[email redacted]

Play in Education, Development and Learning (PEDAL)

University of Cambridge

Memorandum of understanding

“Stepping Stones”

I (class teacher) have read and understood that:

- I will be invited to take part in a one-day workshop and at least three half-day community of practice sessions over 10 weeks.
- Part of these will be audio-recorded.
- I will be encouraged to reflect on my practice and try new things in my classroom, but the final decision as to what I will implement rests with me.
- A researcher will visit my class three times during the project to get a sense of what I am using most. During these visits, the researcher will also ask me some questions about my classroom and practice. This conversation will be audio-recorded.
- I can withdraw from participating from the project at any time or decline to take part in any activity, without giving a reason.

Teacher Name:	
Teacher Signature:	Date:

I (researcher) will guarantee that:

- I will consider the *Ethical Guidelines for Educational Research* the British Education Research Association (BERA) and the *Code of Ethics and Conduct* of the British Psychological Society (BPS) for conducting this research study
- I will abide by GDPR data protection legislation in the collection and storage of these research data, in particular:
 - Data will be kept confidential in password-protected files. If data are shared for re-use by other researchers, I will ensure that they also abide by GDPR legislation in their handling of the data.
 - Data files will be anonymised (e.g. names will be replaced with codes or pseudonyms so individuals cannot be identified).
- Research participants can withdraw without giving a reason at any time.

Researcher Name:	
Researcher Signature:	Date:

Appendix B - Process for Translating the Theory of Change into Actionable Strategies

Whilst the process of translating research findings into actionable practice has been continuous, there were key moments that led to the final selection of 11 strategies.

In November 2018, we reviewed different pedagogies and programmes for the ways in which they provided choice and challenge, as well as the role of the teacher and the environment. These were not literature reviews, but rather reflections based on existing knowledge within the team and further reading. The pedagogies and programmes we considered were: Montessori, High/Scope, the Chicago School Readiness Program (CSRP), Tools of the Mind, Reggio Emilia, Steiner/Waldorf, project-based and inquiry-based learning, Freinet, and Mantle of the Expert.

Early in 2019, we worked to clarify what we meant by *choice* and *challenge*, and wrote definitions as well as operationalisations of each. This was accompanied by further reviews to glean examples from existing research and published pedagogical texts. Drawing from this literature, we created documents to exemplify what we meant by *choice* and *challenge* as well as anticipating ways in which the terms could be misinterpreted or enacted in less meaningful ways. We also began to distinguish between different levels at which these might be operationalised (i.e. classroom organisation, curriculum & planning, lessons & tasks, and ‘in-the-moment’ interactions). Table B1 is an example of such a document for ‘choice’ at the ‘classroom organisation’ level. Whilst the list was not intended to be exhaustive, it was an attempt to compile, based on existing research and pedagogies, all the ways we could find in which children could exercise meaningful choice and experience appropriate challenge. Therefore the process was more akin to collecting and collating practices, rather than any sort of systematic review.

Table B1 *Examples of What Choice (from More Meaningful to Less Meaningful) and No Choice Could Look Like at the Classroom Organisation Level.*

	Meaningful	Somewhat meaningful	Not very meaningful	No choice
Choosing where to sit	Children are free to sit where they like during activities, as long as they don’t disturb the class’s learning. When children make choices the teacher feels are inappropriate (too far to	Children are given the choice to stay at their table or come to the carpet	Children name friends they would like to sit with, which is used to inform a seating plan.	The teacher creates a seating plan.

	hear, next to someone who distracts them), the teacher provides the child with a chance to reflect and find somewhere else to work.		Children choose where they sit for reading time. They then keep this spot for 2 weeks. The order in which they choose is randomly decided.	
Choosing how long to do a task for	Children are given long periods of time in which they can choose when they will do the tasks they have to complete, so they control the pace and the amount of time they spend on each. The teacher might provide children with support if they are at risk of not completing all their tasks, or discuss postponing when the task should be completed.	Children are given a set amount of time, but are given an opportunity to continue beyond that time if they wish to.	Children are asked how long they would like to work for, then a timer is set. When the time has elapsed, children are asked to stop.	Children are given a set amount of time, and frequently reminded.
Choosing what task to do	The classroom is organised so that tasks/games are freely accessible. Children can choose whichever task they want to do, as long as they put it back when they are done.	Children are given a folder (or tray, or box) of tasks, e.g. handwriting, spelling practice. They are given a period of time during which they can choose which task they do from the folder.	Children are given a choice of two or three versions of a task (e.g. with an 'instructionally irrelevant' theme, or different problems).	Children are given a set task.
Choosing who to work with	Children can choose who they work with, whether it is alone, with a peer or in a group. They can choose who they ask for support, they can leave or join groups unless it would disrupt the learning. They are helped to reflect on who might be the best person to work with for a particular task, rather than always choosing their friend.	Children can choose who they work with on a task.		Children have to sit according to a seating plan made by the teacher.
Choosing when to move around the classroom	Children are free to move around the classroom as they wish as long as they don't disturb the learning of others.	Children are allowed to move around the classroom for specific purposes, e.g. get resources, work with a peer.	Some children are given the right to move around the classroom as a special privilege (?)	Children are not allowed to move from their seat unless given permission or asked to by the teacher. Transitions follow a routine (e.g. counting on fingers 1, 2, 3) so children's movements are regulated by the teacher.
Choosing how to sit	Children are allowed to sit in whatever way is comfortable (as long as it's safe). The classroom is organised so that this does not disrupt learning (e.g. consideration for noise, blocking views of the board, etc.).			Children have to sit with their legs crossed ('criss cross applesauce'), with their hands on their lap, or if on a chair with their back straight, their chair tucked in and their feet on the floor.

Choosing what materials to use	The classroom is set up so that children know where resources are and how to put them back when they are done. The teacher involves them in conversations about which materials are appropriate for different purposes.	When children express the wish to use specific materials, they are allowed to. If it is not appropriate, the teacher gives a rationale for not allowing the child to use it.	The children are given a choice of options of materials to use.	Children are given set materials to use.
--------------------------------	---	--	---	--

Whilst I initially carried out this work on *choice*, a research assistant helped with making parallel documents for *challenge*. We then compiled these into a single document which highlighted what layer the practices belonged to (e.g. classroom organisation), and whether they were more related to *choice* or *challenge*.

We also fleshed out what each of these strategies meant and began to collect extracts of real practice to exemplify those (which would later become our ‘gold standard’ examples). However, the very large number of practices would have been overwhelming for teachers new to these practices. Whilst we eventually shared the full table with the teachers towards the end of the Stepping Stones programme as ‘further reading’, the team decided to narrow down the core list to 11 strategies, of which three were highlighted at the very start of the programme (see Appendix C). These were chosen to provide a breadth of different layers and both choice and challenge. In addition, they were strategies which teachers could focus on one at a time. Finally, through further discussions within the team we grouped these practices according to features and goals they shared (e.g. problem solving), though the same practice could feature in different groups.

This is the point the work had reached when we first tested to Stepping Stones programme with teachers. However, this work has since then been continued, with the strategies and the groupings changed and refined over time.

Table B2 *Table of Compiled Practices for Choice and Challenge*

Groups of practices	Specific practices	Layer of practice				Choice	Challenge
		Organisation, systems and structures	Curriculum and planning	Lesson and task	In the moment interactions		
Assessment for learning	Use elaborative utterances to challenge students to think more deeply about their response (*)						X

	Follow up children's answers with questions that vary in specificity and difficulty depending on the amount of support students require (*)						X
	Maintain contact with parents regarding their child (*)					X	X
	Help children make appropriate choices by monitoring the efficacy of their choices					X	X
	Stand back to observe, listen and notice children's engagement and learning					X	X
Children evaluating their work	Have a closure routine to help children evaluate their work and plan what happens next					X	X
	Involve children in evaluating successful examples, strategies and each other's work					X	X
Opportunities for problem solving	Plan open-ended tasks					X	X
	Plan to provide strategies for children to independently solve problems they may encounter					X	X
	Allow children to use a preferred method for doing a task					X	
Control over level of difficulty by the child	Ask open-ended questions (*)					X	X
	Allow children to choose how to present their work					X	X
	Provide activities that are 'low threshold, high ceiling'					X	X
	Allow children to choose the topic for their work (within the context of the learning e.g more or less novel or complex topics) (*)					X	X
Extending thinking	Ask open-ended questions (*)					X	X
	Use elaborative utterance to challenge						X

	students to think more deeply about their response (*)						
	Allowing divergent opinions					X	
Children choosing their own task	Organise the classroom so that tasks/games are freely accessible. Children can choose whichever task they want to do, as long as they put it back when they are done.					X	
	Allow children to choose how to present their work (*)					X	X
Children choose where to work	Allow children to move around the classroom as they wish as long as they don't disturb the learning of others					X	
	Allow children to sit where they like during activities, as long as they don't disturb the class' learning					X	
	Allow children to move to a location free of distractions, where there is more space or by moving closer so they can hear/see more easily					X	X
Children choose who to work with	Have a system whereby children can choose who to work with					X	
	Have a system whereby students share strategies before asking the teacher for support						X
Children choosing materials	Involve students in conversations about which materials to use for different purposes						X
	Set up the classroom so that children know where resources are, how to access them and how to put them back when they are done					X	X
Children choosing how long to work on a	Plan long, uninterrupted periods of time in which children can set their own goals and control the pace and the					X	X

	amount of time they spend on each task						
	Provide children with support if they are at risk of not completing their goals, or discuss when the task should be completed						X
	Encourage and support children to maintain focus even when faced with distractions						X
Children's interests	Maintain contact with parents regarding their child's learning (*)					X	X
	Provide children with a rich context, hook or stimulus where their questions might emerge - and use these to plan					X	
	Allow children to choose the topic for their work (e.g. more or less novel or complex topics) (*)					X	X
	Involve children in deciding the purpose and goals for their learning					X	X
	Respond to children's interests as they emerge by being flexible in planning lessons and by drawing on these interests as learning opportunities					X	
Sensitivity	Recognise and verbalise children's thoughts and feelings (H)						X
	Actively respond to children's cues of distress						X
	Model and foster empathic behaviour for and among children						X
	Encourage children by demonstrating your interest in their work						X
	Allow divergent opinions					X	
	Use informational rather than directive language (*) (S)					X	

Scaffolding	Anticipate and adjust the level of support depending on your understanding of the child						X
	Follow up children's answers with questions that vary in specificity and difficulty depending on the amount of support students require (*)						X
	Using informational rather than directive language (*)					X	
	Help children make appropriate choices by helping them think metacognitively (S)					X	X

Note. Strategies marked with a star (*) occur more than once as they can belong to different groupings.

Appendix C - Stepping Stones Programme Strategies to Support Children's Autonomy

Appendix C1 - First Three Strategies Introduced in Workshop 1

Ask **open-ended** questions

Open-ended questions are questions with no single or 'right' answer. They invite the children to provide their own perspective and ideas, and can be used to probe children's knowledge and understanding, extend their reasoning and encourage them to reflect on themselves as learners. Open-ended questions can be about the process of learning (e.g. 'How could we do this differently?') or about subject-specific knowledge (e.g. 'Where do we come from?'). This can open up a discussion and enable teachers to assess children's understanding, including potential misconceptions.

Example:

"Dawn moved around the room observing, listening and occasionally joining in conversations. She frequently provoked the children's thinking by pondering tentatively aloud using 'what if' framing [...], for example: I was wondering what would happen if you used the material differently? and It might have been a challenge to move soldiers from the battlefields to the hospital. What kinds of ways might they have achieved that I wonder? and Would it help you at all to make a floor plan? - I suppose it might take too much time? What if you did this differently-what options can you think of? (transcript, p.3) Her speculative stance indicated a genuine interest in the children's ideas and her language helped them maintain a sense of agency and influence over their work."

Source: Cremin, Burnard, & Craft (2006). Pedagogy and possibility thinking in the early years. *Thinking Skills and Creativity*, 1(2), 108-119.

Respond to children's interests as they emerge by being **flexible** in planning lessons and by drawing on these **interests** as learning opportunities.

Children's sense of ownership over their learning is greater if the subject matters to them, and when they are curious or interested in a topic. Teachers can harness children's enthusiasm when opportunities arise (sometimes unexpectedly) by being responsive in their teaching.

Example:

"One day, during a group literacy lesson, six-year-old Mary discovered a picture of a volcano and told Ms Bailey and the class that she wanted to build one. Her excitement spread, and soon the whole class wanted to figure out how to make a volcano. Ms. Bailey decided to put her lesson plan aside and asked the children what they knew about volcanoes. Then she asked the children to make a model of what their volcano might look like and write a list of materials they would need to make it. Soon, Mary had a group of four girls working intently with her and sharing ideas. The girls drew their models on paper and then labeled the different parts. Students wandered around the class asking their peers about their drawings and gathering ideas. Back in their group, Mary and the other girls could be heard laughing, asking one another questions about their drawings, describing and defending their work, and excitedly telling each other what they had seen at other tables. At [break time] over the next couple of days, they enacted their volcano-making ideas with models made of dirt, rocks and sticks. [...]"

On Monday, Mary arrived at school with a large clay model of her volcano complete with [...] dish detergent, baking soda, vinegar and food dye. Immediately the children gathered around Mary and her volcano, asking questions and begging to be the one to put the ingredients in."

Source: Adair (2014). Agency and Expanding Capabilities in Early Grade Classrooms: What It Could Mean for Young Children. *Harvard Educational Review*, 84(2), 217-242.

Use of **elaborative utterances**

Elaborating on children's questions and responses during a discussion can include rephrasing their responses, adding new pieces of information or asking questions to help children justify their answer or their methods. These elaborations encourage children to think more deeply about their initial response.

Example:

Adam: "If they didn't have roots it would break."

Susan (teacher): "What might happen to the plant then?"

Jake: "Then the plant would probably die."

Susan: "Okay...could you tell me what's the purpose of the roots?"

Jake: "It's kinda like the heart because if the plant didn't have roots it wouldn't live, and with us, if we didn't have hearts we wouldn't live and stuff, so the plant needs the roots just like we need our heart."

Susan: "And what's the purpose of our heart?"

Jake: "The purpose of our heart is to run blood flows through our body."

Susan: "If you're saying that the roots are like the heart, which is very important to us, right, and you're saying the roots are very important to the plant, what's the purpose of the roots?"

Jake: "The purpose of the roots is to keep the plant alive, just like the purpose of our heart."

Susan: "Okay, be specific! How does it keep the plant alive?"

Rose: "Because it sucks up the water."

Blair: "So one of the purposes, or one of the functions of the roots is to pull water into the plant".

Source: Chen, Hand, & Norton-Meier (2017). Teacher Roles of Questioning in Early Elementary Science Classrooms: A Framework Promoting Student Cognitive Complexities in Argumentation. *Research in Science Education*, 47(2), 373-405.

Appendix C2 - Further Strategy Cards Introduced in Workshop 2

Observe, listen and notice children's engagement and learning

This practice allows teachers to notice children's interests, strengths and the work they produce.

What does it look like in the classroom?

- Look for signs of engagement e.g. facial expressions, questions, conversations with peers
- Take note of topics and associated resources
- Notice how long a child focuses on an activity

Why does it work?

This helps teachers to anticipate the choices a child will make and areas where they would work well with challenge.

'In the moment' Interactions	Lesson and Task	Curriculum and Planning	Classroom Organisation, System and Structures
------------------------------	-----------------	-------------------------	---

Continuous provision or the 'prepared environment'

This is a system where children choose from a range of activities or resources set out by the teacher.

What does it look like in the classroom?

- Make activities and resources freely accessible for children to choose
- Offer extended periods of time for children to engage with activities
- Plan activities carefully to ensure deep and varied learning experiences

Why does it work?

Choosing an activity that is personally interesting helps to increase intrinsic motivation. Extended periods of time mean learning can be deep and personalised, and children have an opportunity to work through challenges.

'In the moment'
Interactions

Lesson and Task

Curriculum and Planning

Classroom Organisation,
System and Structures

Open-Ended Tasks

These tasks have multiple solutions, multiple ways to achieve the solution, or no solution. The purpose may be to generate and evaluate ideas.

What does it look like in the classroom?

- Discuss a moral or philosophical question
- Set up an open investigation of an intriguing phenomenon or observation
- Tackle complex tasks, where children explore an issue or topic over time

Why does it work?

Children use their executive functions whilst making and reviewing their decisions. The freedom to explore increases their feelings of autonomy.

'In the moment'
Interactions

Lesson and Task

Curriculum and Planning

Classroom Organisation,
System and Structures

Evaluate successful strategies & each other's work

Children explain and justify their methods, and reflect on those used by other students.

What does it look like in the classroom?

- Children are asked to choose a method and explain their choice
- Comparing alternative solutions to a problem: which one works best, and why?
- Children discuss pros and cons of different approaches as a class

Why does it work?

Children are challenged to reflect on their choices. This will help become better at choosing an appropriate strategy and recognise when they need to try something different.

'In the moment'
Interactions

Lesson and Task

Curriculum and Planning

Classroom Organisation,
System and Structures

Provide strategies for independent problem solving

Explicitly teaching children how to solve a problem, or drawing children's attention to the strategies they use.

What does it look like in the classroom?

- Avoid immediately correcting mistakes or providing the right answer
- Ask, 'What could you do?' and encourage children to reflect on their first attempt
- Provide and model strategies for solving the problem

Why does it work?

Children understand the learning process and can draw on this knowledge while working autonomously. It empowers them and helps them to make good decisions about which strategies to use.

'In the moment'
Interactions

Lesson and Task

Curriculum and Planning

Classroom Organisation,
System and Structures

Model and encourage empathy among children

Demonstrating to the children that you understand how they are feeling and putting it into words for them.

What does it look like in the classroom?

- Recognise and verbalise children's emotions i.e. model appropriate concern when a child is hurt or distressed, and show an interest when they are engaged and excited.
- Encourage children to communicate with each other about their emotions
- Use books, songs and art to spark discussion about emotions

Why does it work?

These practices help children to understand their emotions. This will help them to self-regulate in the future, if their emotions (or those of their peers) interrupt learning.

'In the moment'
Interactions

Lesson and Task

Curriculum and Planning

Classroom Organisation,
System and Structures

Allow Different Opinions

Using unexpected answers as learning opportunities, instead of refocusing children to the expected answer.

What does it look like in the classroom?

- Respond positively to a range of answers
- Explore incorrect answers with questions and encourage children to justify their answers
- Nurture a 'no wrong answer' environment when sharing ideas, and encourage children to be accepting of others ideas

Why does it work?

Teachers can challenge children's misconceptions when asking questions about their reasoning. Allowing children space to think in their preferred way also increases children's autonomy.

'In the moment'
Interactions

Lesson and Task

Curriculum and Planning

Classroom Organisation,
System and Structures

Involve children in deciding their learning goals

Children help to shape the direction of their learning. Teachers build upon children's preferences and interests when involving them in choosing the purpose of their work.

What does it look like in the classroom?

- Children are allowed to make choices about the purpose of their work
- A teacher might propose activities having noticed a child's interest, and lets them take it up if they wish to.
- Children are given contexts so that purposes might emerge

Why does it work?

Children are more likely to persist, enjoy the learning, and be interested in the outcome of the task when they have a purpose that is meaningful to them.

'In the moment'
Interactions

Lesson and Task

Curriculum and Planning

Classroom Organisation,
System and Structures

Appendix D - Overview of Content for the Stepping Stones Community of Practice Workshops

Workshop 1 (Focus: Introduction of key ideas)

Introduction to the Project

Ice Breaker

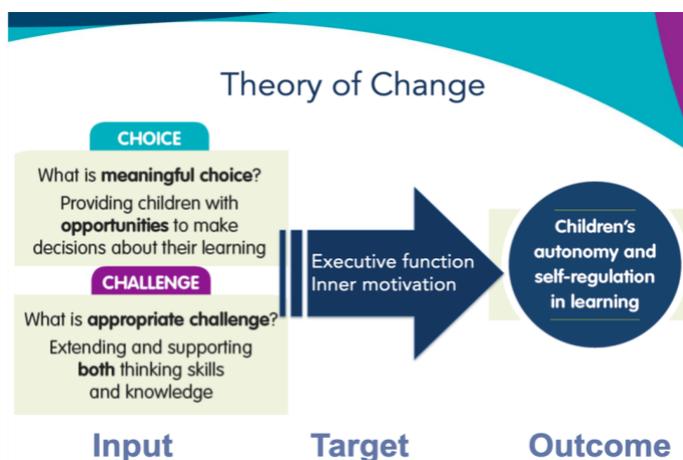
Introduction to the Project and Rationale. Presentation of the team and the Stepping Stones project, including past work.

Joint Agreement - Memorandum of Understanding (MoU). Researchers led and facilitated a discussion around a shared vision for the Community of Practice and ‘ground rules’, e.g. openness, non-judgement. Researchers introduced the notion of ‘teachers as co-researchers’ and gave opportunities for teachers to share their own needs and expectations for the workshops.

Introduction to the Theory

Researchers presented background research in the context of choice and challenge, as well as previous work with teachers, the Theory of Change (Figure D1), and a brief introduction to executive function and inner motivation. This also included a presentation of *choice* and *challenge* as the focus inputs for the present Community of Practice.

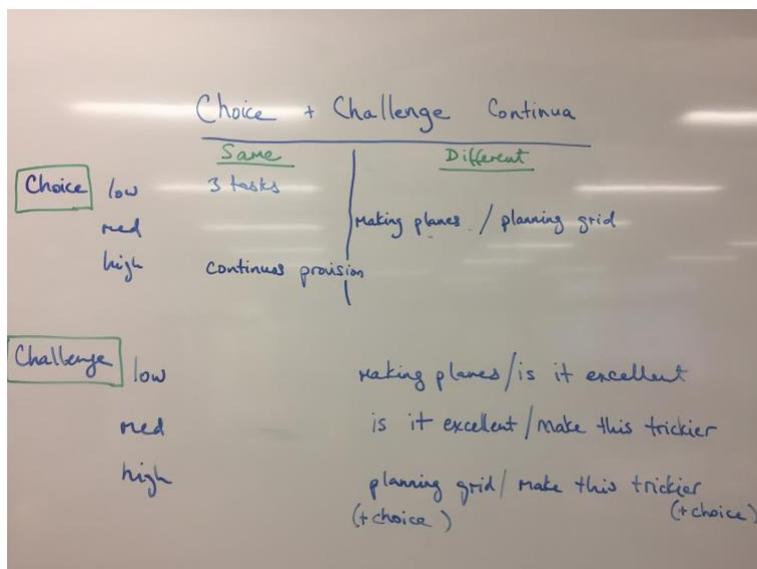
Figure D1 *Stepping Stones Theory of Change in March 2019*



Putting the Theory into Practice

Discussion of Choice and Challenge Examples. Teachers were given examples of choice and challenge, including examples less likely to foster autonomy in children, in order to provoke discussion and reflection with regards to classroom practice. Teachers placed these examples along a continuum of choice and of challenge (Figure D2). Teachers discussed this in small groups then fed back to the whole group, and each group's ideas were summarised in a table.

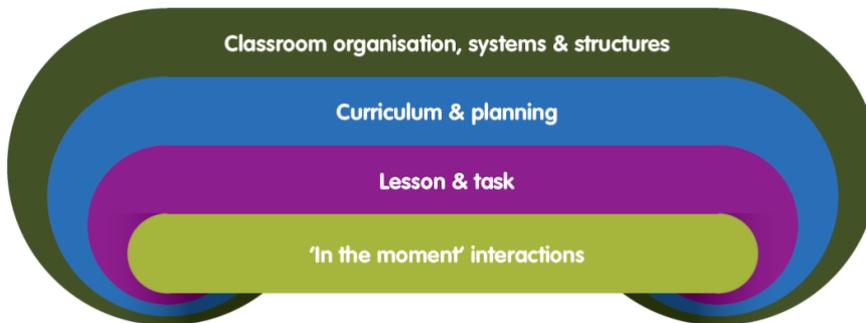
Figure D2 Choice + Challenge Continua Activity Summary



Choice & Challenge 'Gold Standards'. Teachers were introduced to examples that are illustrative of practice thought to promote autonomy in children taken from the research literature and pedagogical texts. These examples related to three strategies for *choice* and *challenge*: flexible planning to follow children's interests, open-ended questions and elaboration (see Appendix C). Teachers labelled texts depending on whether they thought them to be illustrative of teachers providing *choice* or *challenge* or both.

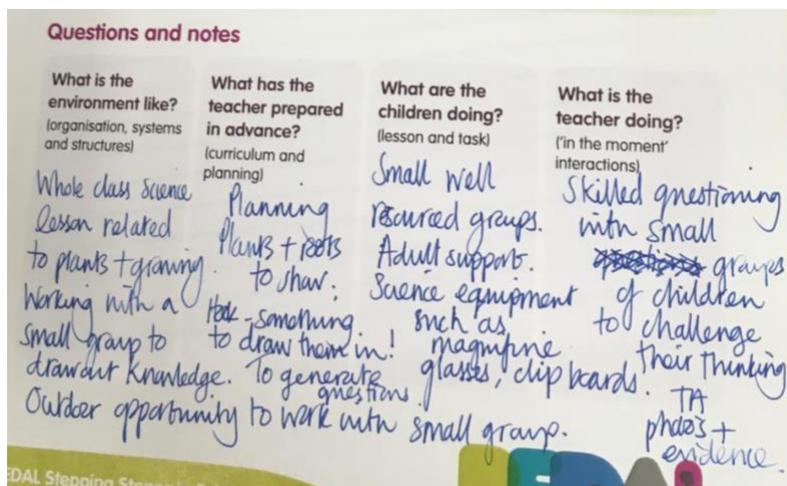
Choice & Challenge in Layers of Practice. Teachers were introduced to the idea of layers of practice, that is to say levels at which the teacher might be able to bring about choice and challenge (Figure D3).

Figure D3 *Layers of Practice*



Teachers were asked to choose one of the three strategies discussed in the previous activity (e.g. one they might be interested in) and to discuss how it might be enacted at the different layers of practice, i.e. what it might look like in practice, using a real classroom scenario if they wished. Teachers were provided with question prompts to help guide their discussion and reflection.

Figure D4 *Layers of Practice Question Prompts and Example Teacher Notes*



Planning

Planning for a Lesson. The research team presented an example of planning a series of lessons using the principles outlined in the workshop, using a flexible enquiry framework in a science lesson. Teachers had been asked to bring a lesson plan or a topic they would use in the coming term. They were given a period of time (about 30 minutes) to plan for the

coming term, using what they had learned in the workshop and drawing on the team and other's input if needed.

Goal Setting & Wrap Up. Teachers were asked to set themselves short-term, achievable goals (“What will you change tomorrow and next week?”), and longer-term, ambitious goals. Teachers completed feedback forms.

Workshop 2 (Focus: Extend)

Reflect

Individual Written Reflection. Teachers spent some time reflecting, and writing if they wished, about what they had tried out since the last workshop. They were given a narrative frame as support. Teachers then took turns sharing something they tried.

Recap

Teachers were given a very brief reminder about the strategies presented in the previous workshop.

Extend

New Strategies. Teachers were introduced to 8 additional strategies to support choice and challenge in the classroom. These were presented as ‘strategy cards’ (Appendix C)

Video Carousel. In order to make sense of the new strategies, teachers visited stations around the room, in pairs, where they could watch different videos and discuss which strategies the teachers might be using, and to what extent they were doing so in a way that supports choice and challenge. Teachers then discussed their thoughts as a group.

Plan

Shared Example. The team shared an example plan from a previous partner school. In small groups, teachers discussed how they would adapt this for the practice they were focusing on.

Planning Time. Teachers were given a period of time (about 30 minutes) to plan next steps. This could be concrete lesson plans, or higher-level objectives and what they would do to realise them. The research team supported teachers, e.g. with questions they had.

Critical Incident Interviews.

Teachers were invited to reflect on a critical incident since the last workshop, i.e. a moment where something went particularly well, or did not go to plan, and which made the teacher learn about their practice.

Wrap Up

Teachers set themselves new goals/ steps until the next community of practice and completed feedback forms. Teachers were told about the platform where the videos were hosted as well as all materials from the workshops, which they could access at any time.

Workshop 3 (Focus: Meeting Challenges)

Reflect

Sharing Experiences. Two teachers were invited to present their experience to the group, including what they had focused on, challenges they encountered, what they did in response, and what contextual factors they had to take into account. Teachers were invited to report on lessons learned and to reflect on their journey so far. Other teachers were invited to ask questions.

Reflecting on Challenges. All teachers were asked to reflect on their experiences and in particular on something they had found challenging. Teachers wrote their ‘challenge’ on post-it notes to stick to the board at the front, which were then grouped together by theme.

Figure D5 *Challenges Board, Organised by Theme by the Research Team*



In each theme, a teacher was selected to explain and expand on what they meant and other teachers who faced similar challenges were invited to respond or build on this.

Recap

Teachers were given a very brief reminder about the strategies presented in the previous workshop and given the opportunity to ask any questions they had.

Plan

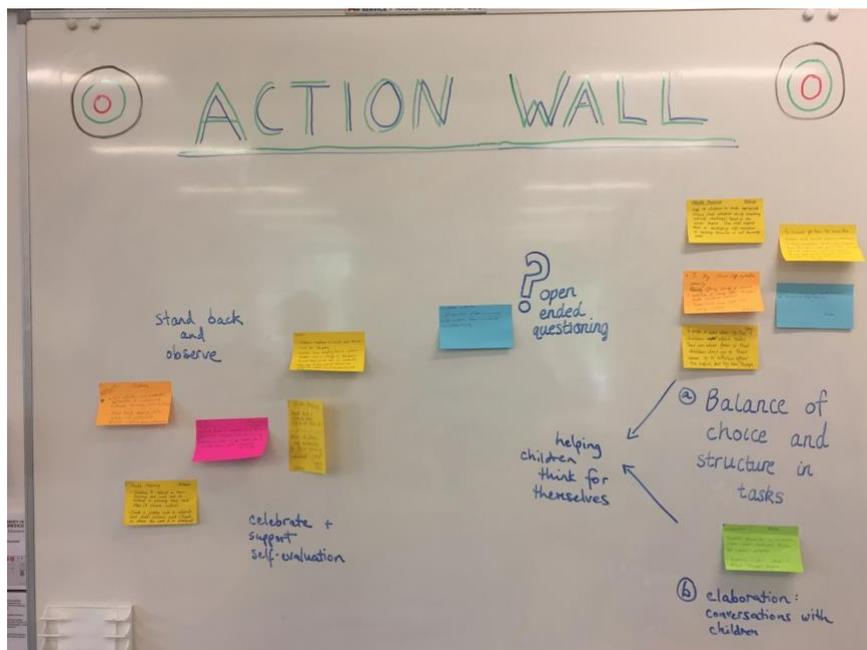
Planning Time. Teachers were given a period of time (about 30 minutes) to plan next steps. This could be concrete lesson plans, or higher level objectives and what they would do to realise them. The research team supported teachers, e.g. with questions they had, and laptops were provided to access resources. This included resources from previous workshops as well as illustrative examples of practice for the 8 strategies introduced in workshop 2.

Wrap Up

Action Wall. Teachers set themselves new goals/ steps for the next community of practice, which they were told they would revisit at the final workshop. Actions were written on a post-it note that was stuck on an ‘Action Wall’.

Teachers completed feedback forms.

Figure D6 Action Wall, Organised by Teachers’ Strategy Focus



Workshop 4 (Focus: Reflect)

Reflect

Reflection on Action Plans. Teachers were given time to reflect, and write using a frame if they wished, on the actions from the previous workshop. Teachers then shared these reflections as a group.

Sharing Experiences. One teacher shared something from their practice since the last workshop e.g. a problem they encountered and the solution they found to the problem. This was followed by a group discussion and inviting other teachers to share if they wished.

Finding Solutions. Teachers were introduced to a table with solution prompts to some of the challenges they mentioned at the previous workshop. These took the form of questions and invitations to explore ideas rather than ‘fixes’. Teachers discussed these in pairs.

Evaluating Choice and Challenge. Teachers were invited to reflect on the question, “How do we know it’s working?”. On large pieces of paper, they reflected on this question in different situations (small group, one-to-one, peer interactions, classroom-level). They were given additional questions to frame their reflection e.g. “What would you expect to see when practices are going well? What does it look like in terms of children’s behaviours, e.g. engagement?”. Teachers annotated the sheets to create ‘success criteria’, to use for reference in future planning, and as a reflection tool.

Wrap Up

Teachers filled in questionnaires about their experience.

Appendix E - Stepping Stones Interview Protocols

Appendix E1 - Initial Interview Protocol

Overview:

- The purpose of the interviews is to understand the implementation of intervention materials from the teacher's perspective. This will provide teachers with an opportunity to share their feelings, goals and insights about the feasibility of the intervention.
- The aim of the initial interview will be to establish the context of the classroom and the teacher's own beliefs.
- The aim of the final interview is to answer the research question: What changes to their practice can teachers implement by taking part in a professional development programme focusing on the provision of 'choice' and 'appropriate challenge'? This includes: what kinds of changes teachers make, how much of a change they are able to make, and what factors, including resources, structures and aspects of the development programme, have affected the ease or difficulty of teachers modifying their practice.

Notes if meeting the teacher for the first time:

- If the memorandum of understanding has not been signed, make sure it has been signed
- Introduce yourself (name, position in the team)
- Ask if you can call the teacher by their first name

General notes:

- As an interviewer, try to speak as little as possible – your main job is to listen and show interest in the stories the teacher tells you
- Try to be sensitive to issues that trigger distrust and to the informant's 'inner voice', body language, silences, laughs, etc.
- Follow your hunches and what the teacher is telling you rather than adhere to the script of questions – but make sure the teacher is telling you something about all areas of inquiry outlined below
- You can make notes to keep track of and focus on what the teacher is saying

- Be polite and show your interest in what informants are saying – but err on the side of formality rather than familiarity
- **Make sure your questions are not leading.** For example, if a teacher does not mention a specific practice, do not ask, ‘What about doing this?’ or ‘Do you do this in your classroom?’. This would bias answers, especially if those questions are not asked consistently.

General procedure:

- Remind the teacher of the purpose of these interviews, e.g. “We want to find out your views and feelings about the project”.
- Tell them the chat should take 30 to 45 minutes
- Remind the teacher you will use an audio recorder:
“As we’d mentioned before, I am going to be recording our conversation so we can use it later. The first thing I will do is state some details including your name. This is so we can identify the audio segment later, but in all transcripts that name will be replaced by a code. The audio file will be kept in a password protected drive and not shared outside of the team.”
- Remember to start the audio recorder and tell the teacher when you do.
- State your name, the date, and the name of the teacher you are interviewing
- If the conversation is still taking place after 40 minutes, let the teacher finish but don’t ask any more questions.
- At the end of the conversation, tell the teacher you are going to stop recording and press stop. What the teacher now says is off the record.
- Thank the teacher for their time.
- After the interview, as soon as the interviewer can find a discreet place to do so, the interviewer should make any notes and reflections on the interview process (e.g. questions that were not followed up on, and reasons why) as well as general observations and reflections (hunches, body language noted, etc.)

Initial interview:

- The focus is on understanding the teacher’s perspective on the practices we will later ask them to try out, as well as on the project.

- Questions relate to four focus areas: Teaching experience and general pedagogy, Teachers' current practice, Teachers' beliefs and practice relating to the intervention materials, Teachers' thoughts and beliefs about the intervention

Focus areas	Possible prompts – underlined questions MUST be asked
Teaching experience, general pedagogy	<u>I'd like to know more about you as a teacher. How did you come to be a teacher?</u> <u>How long have you been a teacher?</u> Who inspires you in the education world?
Teachers' current practice	How would you describe your school? What is it like to teach there? I'd like to know about what a typical day is like for you. Talk me through a day in your classroom. What is your student's motivation like? What do you find helps to motivate them? If you had a magic wand to change one thing about what happens in your classroom, what would you change?
Teachers' beliefs and practice relating to the intervention materials	What are your views on giving children more choice in the classroom? What kinds of choices do you give your students? What kinds of decisions do your students make about what they will learn? What strategies do you use to help children access the learning at the right level of challenge for them?
Teachers' thoughts and beliefs about the intervention	How did you become involved in the project? How was the decision made? How do you feel about taking part in the project?

Appendix E2 - Critical Incident Interview Protocol (as written by Adams, 2019)

General Points to Remember

These are semi-structured interviews, and therefore we as interviewers have license to ask follow-up questions to probe further into participants responses, where this might elucidate richer data. For example:

- Asking for more details
- Asking for clarification
- Asking for an example

- As far as possible, try to remain passive and neutral during the interview. Take a non-judgemental stance and pay attention to your verbal and non-verbal cues to ensure these are aligned with such a stance.
- Tolerate silence, particularly when the teacher is attempting to recall the specific incident they wish to discuss.
- However, also remain attentive to what the teacher is saying and use prompts where necessary to keep the conversation in line with the questions asked.
- Use checks to clarify any ambiguities in what the teachers are saying.

Questions/Procedure

1. Initial briefing (full group), to cover:
 - a) Aims of the research (i.e. investigating feasibility of using teaching practices)
 - b) Logistics of how the interviews will be conducted.
 - c) Reiterating that interviews are not an assessment of teaching practice, but an exploration of how feasible these practices were for the teachers.
 - d) Audio recording of the interviews.
 - e) Confidentiality of responses.

2. Transition to individual interview room; ensure researcher and interviewer are sitting at a table, at a 90° angle to each other; start audio recording.

3. Preamble (individual)

We would find it really helpful to hear about a time, since the last Community of Practice, when you tried out one of the teaching practices that was introduced.

It would be great if we could focus on one particular time that stands out to you as important, in terms of how easy or difficult you found it to use the teaching practice. Can you think of one?

Pause to allow teacher to think of a specific situation. When appropriate, continue to Question 1.

Question 1: Can you describe the situation that comes to mind?

Prompts for Question 1 (if necessary):

If the teacher begins to talk generally about trying out the teaching practice(s), encourage them to focus upon a specific instance, for example using the following prompts:

- Could you describe a particular time when that happened that stands out to you as important, in terms of how easy or difficult you found it?

If the teacher has not tried out any of the teaching practices, then probe the reasons for this, keeping the issue of feasibility in mind. For example, you could use prompts such as:

- Were there any particular reasons why you did not use the teaching practices?

Question 2: How did you find this experience, in terms of how easy or difficult it was to try out the teaching practice?

Prompts for Question 2 (if necessary):

If the teacher only talks about their experience of using the teaching practice without discussing the reasons why, encourage them to also explain these reasons, for example using the following prompts:

- Did anything make it easy to try out this teaching practice?
- Did anything make it difficult to try out this teaching practice?

Appendix E3 - Final Interview Protocol

Overview:

- The purpose of the interviews is to understand the implementation of intervention materials from the teacher's perspective. This will provide teachers with an opportunity to share their feelings, goals and insights about the feasibility of the intervention.
- The aim of the final interview is to answer the research question: What changes to their practice can teachers implement by taking part in a professional development programme focusing on the provision of 'choice' and 'appropriate challenge'? This includes: what kinds of changes teachers make, how much of a change they are able to

make, and what factors, including resources, structures and aspects of the development programme, have affected the ease or difficulty of teachers modifying their practice.

- The focus is on getting concrete answers to the research questions, but drawing on the teachers' interpretations of what happened in their classroom throughout the project
- There are three focus areas which reflect the three research (sub)questions: kinds of changes to practice, amount of change to practice and barriers and affordances for change.

General notes:

- As an interviewer, try to speak as little as possible – your main job is to listen and show interest in the stories the teacher tells you
- Try to be sensitive to issues that trigger distrust and to the informant's 'inner voice', body language, silences, laughs, etc.
- Follow your hunches and what the teacher is telling you rather than adhere to the script of questions – but make sure the teacher is telling you something about all areas of inquiry outlined below
- You can make notes to keep track of and focus on what the teacher is saying
- Be polite and show your interest in what informants are saying – but err on the side of formality rather than familiarity
- **Make sure your questions are not leading.** For example, if a teacher does not mention a specific practice, do not ask, 'What about doing this?' or 'Do you do this in your classroom?'. This would bias answers, especially if those questions are not asked consistently.

General procedure:

- Remind the teacher of the purpose of these interviews, e.g. "We want to find out your views and feelings about the project".
- Tell them the chat should take 30 minutes to an hour
- Remind the teacher you will use an audio recorder:

"As we'd mentioned before, I am going to be recording our conversation so we can use it later. The first thing I will do is state some details including your name. This is so we can identify the audio segment later, but in all transcripts that name will be replaced by a code. The audio file will be kept in a password protected drive and not shared outside of the team."

- Remember to start the audio recorder and tell the teacher when you do.
- State your name, the date, and the name of the teacher you are interviewing
- Check at 30 minutes how far into the question schedule you have got - leave at least 20 minutes for the tasks.
- If the conversation is still taking place after 55 minutes, let the teacher finish but don't ask any more follow-up questions and move on to the final question.
- At the end of the conversation, tell the teacher you are going to stop recording and press stop. What the teacher now says is off the record.
- Thank the teacher for their time and for their participation in the project.
- Check their contact details for future updates and debriefs.
- After the interview, as soon as the interviewer can find a discreet place to do so, the interviewer should make any notes and reflections on the interview process (e.g. questions that were not followed up on, and reasons why) as well as general observations and reflections (hunches, body language noted, etc.)

Focus areas	Possible prompts
Kinds of changes to practice (RQ 2.1)	<p><u>I'd like to know about the changes you made to your practice over time – can you talk me through the different things you tried throughout the project?</u></p> <p><i>Why did you decide to try these in particular?</i></p> <p><i>What kinds of changes have you kept? Which did you stop doing? Did you change the way you implemented any of these practices?</i></p> <p><u>Do you plan on carrying on with these practices once the project is finished?</u></p>
Amount of change to practice (RQ 2.2)	<p><u>How often do you use these practices/ systems?³</u></p> <p><i>Do you use them in all subjects or some specific ones?</i></p> <p><i>How regularly do you use these practices?</i></p> <p><i>Do you use these throughout the day or only in some circumstances?</i></p> <p><u>Did this change over the course of the project?</u></p>
Change in conceptualisation of choice and challenge (RQ 2.1)	<p><u>At the beginning of the Community of Practice sessions, we introduced the notions of choice and challenge.</u></p> <p><u>What are your thoughts on choice and challenge now?</u></p> <p><i>Could you explain to me what sorts of choice your children have in their learning now?</i></p>

³ Only refer to practices the teachers mentioned themselves – do not mention or suggest additional practices, unless there is a very important reason to do so, e.g. you notice a discrepancy between what the teacher says and what you saw

Could you explain to me how you challenge children in their learning now?

Has the way you think about choice and challenge changed?

Looking at the whole of the practices we have suggested:

How did you initially feel about these practices?

Were there any practices you did not want try?

Do Q-sort activity – take a picture!

What are your thoughts and feelings on these practices now?

Were there any practices you wanted to try but could not?

Is there anything you tried and decided to stop doing?

Do barrier ranking activity – take a picture!

If applicable: you mentioned you are planning on continuing with some practices. What makes them attractive to you?

Is there anything that might make it difficult for other teachers to make the same changes you made?

Did you benefit from any support other than the communities of practice?

Focusing now on your experience of taking part in the project:

What was it like for you to make changes to your practice?

What about making changes to your practice did you find more difficult?

What about making changes to your practice did you find easier?

How did it feel?

Finally, could you share one good thing that came out of your participation in the community of practice?

Barriers and affordances for change (RQ 2.3)

Q sort activity procedure:

- Tell the teacher you would now like them to do an activity about teaching practices.
- Present the cards with practices to the teachers and remind them these are the practices that were introduced throughout the project.
- Present the blank grid for teachers to fill in. Say “For this task, I would like you to put the practices in the boxes according to how easy or difficult you think it is to implement each of them. This end means it is easiest to do, the other end means it is the hardest to do.” Demonstrate with the surfing, skiing and cycling cards. Explain that they cannot put more practices in any column than there are boxes. Again, demonstrate with the surfing, skiing and cycling cards.
- Ask teachers if they have any questions and clarify anything they are unsure about regarding the procedure.

- During the task, if they are unsure where to put a card, ask them to focus on whether it is more or less difficult than the other practices on the grid.
- Ask teachers if they want to make any final changes before you move on.
- **Take a picture of the final grid!**

Remember!

- **The conversation is as important as where the cards end up!**
- **Encourage the teacher to think out loud.**
- **Make sure that when they talk about a card, the practice is named (by them or by you) so that a listener can easily understand which practice is being talked about.**

Some questions to guide the conversation:

- What makes this practice easier than others?
- What makes this practice more difficult than others?

Barrier ranking procedure:

- Tell the teacher you would now like them to do an activity about the barriers to implementing these practices.
- Present the barrier cards and explain these were the barriers that came up during the meeting, and ask if there are any barriers or challenges they encountered which are missing. Give them an opportunity to write them on blank cards.
- Ask them to rank them in order according to how much of an issue this was for them, from most relevant them, to least relevant. If teacher feels some barriers are equally important, they may put them together in a row. The idea is to get at how much of a problem it was in their context (as opposed to how much of a problem it might be in general). For example, a teacher might recognise that a lack of support from SLT is a serious barrier, but their own SLT is very supportive so it would be ranked low.
- If one of the labels is in fact not a barrier at all for that teacher, they may leave it out of the ranking entirely.
- Provide as many strips as the teacher feels are needed to add barriers not already given.

- Teachers may wish to organise the strips of paper slightly differently (e.g. leave some gaps between some practices because they are further apart in how important they are to them) – this is fine too as long as they discuss their reasons.
- Ask teachers if they want to make any final changes before you move on.
- **Take a picture of the final ranking!**
- When you are able to, fill in the table giving the rank of each barrier (1 for most important issue, 2 for second most important, etc.)

Appendix E4 - Strategies Q-sort Activity

See final interview protocol for instructions given to teachers.

Strategy Labels.

Open-ended tasks

These tasks have multiple answers, multiple ways to achieve the task, or no right or wrong answer. The purpose may be to generate and evaluate ideas.

Allowing different opinions

Using unexpected answers as learning opportunities, instead of refocusing children to the expected answer.

Continuous provision or the ‘prepared environment’

This is a system where children choose from a range of activities or resources set out by the teacher.

Evaluate successful strategies and each other’s work

Children explain and justify their methods, and reflect on those used by other students.

Provide strategies for independent problem solving

Explicitly teaching children how to solve a problem, or drawing children’s attention to the strategies they use.

Involve children in deciding their learning goals

Children help to shape the direction of their learning. Teachers build upon children’s preferences and interests when involving them in choosing the purpose of their work.

Model and encourage empathy among children

Demonstrating to the children that you understand how they are feeling and putting it into words for them.

Observe, Listen and notice engagement

Ask open-ended questions

These are questions with no single or 'right' answer. They invite the children to provide their own perspective and ideas.

Respond to children's interests as they emerge by being flexible in planning lessons and by drawing on these interests as learning opportunities.

Use elaborative utterances

This can include rephrasing their responses, adding new pieces of information or asking questions to help children justify their answer or their methods.

Q-sort Blank Table.

Most difficult

Easiest

Appendix E5 - List of Barriers for the Barrier Ranking Activity

See final interview protocol for instructions given to teachers.

Note: teachers were also given blank slips to add barriers and could discard barriers that were not relevant to them.

- Curriculum pressures
- Children's response to activities
- Children's inability to cope with more independence
- School timetables
- Teaching Assistants not having the right skills
- Lack of resources
- Workload

Appendix F - Running Record Protocol and Template

Appendix F1 - Running Record Protocol

Overview:

- Running records are a way to record naturalistic observations. They are in-the-moment, thick descriptions of events as they unfold. These descriptions should be as objective as possible, though this is limited by where the attention of the recorder is focused. Importantly, the language used is descriptive and not evaluative.
- The running records will be carried out at three time points during the intervention (beginning, middle, end).
- At the end of each session, a debrief will also be organised with the teacher at a time convenient to them. The purpose is not to give feedback but to provide the teacher with an opportunity to share their own perspective on the lesson, such as their intentions when planning the lesson, what they wanted to achieve, and anything they noticed or acted upon which they wish to share.
- The running records will be used to observe any changes to the teachers' practice. In particular, it will allow me to describe the amount of time and the kinds of activities where the teacher actively tries to incorporate greater choice and challenge, compared to the amount of time and kinds of activities spent in 'business-as-usual' teaching.

General notes:

- These are class visits, not 'observations' (negative connotations)
- Aim to see one science lesson and one Literacy or Numeracy lesson – try to keep the types of lessons consistent at each observation point (e.g. a Literacy lesson at each time point, rather than some Literacy and some Numeracy).

Notes if meeting the teacher for the first time:

- If the memorandum of understanding has not been signed, make sure it has been signed
- Introduce yourself (name, position in the team)

Before the start of the day:

- Confirm the lessons to be seen (try not to say 'observe')

- Get a timetable for the day with approximate times, including break/lunch times
- Ask where you can sit so you won't disturb the lesson
- Arrange a time for debriefing that is convenient for the teacher

Running record procedure

(a) Description:

The running record recording sheet (below) has three sections. The first provides space to record (a) whose classroom is being observed, in what school, and at what grade level, (b) who is observing, (c) the date of the observation, and (d) the nature and duration of the activities in which teachers and students are engaged during that observation (e.g., selecting a topic to research, writing the start to a fairy tale). The second section provides space to keep a running record of "what was going on," including verbatim samples of teachers' and students' speech. Although the focus is on choices and support provided by teachers,

The third section lists categories of teaching practice related to choice and structure derived from the literature review. This list of categories provides a conceptual framework for observing in classrooms (and later coding those observations). However, observers are encouraged to refine and expand these categories through their observations.

(b) Observation and records:

Before the observation:

- Fill in the first section with details of the lesson

During the observation:

- position yourself so that you can clearly see and hear the teacher and students without being intrusive.
- record in the second section:
 - o events and actions
 - o a list of times related to events and actions
 - o as much as possible, verbatim speech in teacher–student and student–student interactions.
- Although the focus is on what the teacher does, and what choices and structure they provide, record interactions with students and between students if they are relevant.

Immediately after the observation:

- Read and annotate running records:
 - o Add details regarding events and actions that you did not have time to record during the observation
 - o Fill in gaps in teachers' speech with paraphrases of what they said.
(Paraphrases are marked with square brackets.)
- Once you are satisfied that the running record is as detailed and accurate as your memories will allow, re-read it.
- Add any missing details to the first section (e.g. description of activities)
- In the third section:
 - o note incidents and examples reflected in the original list of analytic categories
 - o record events and actions that suggest refinements or additions to those categories (i.e., emerging categories).

(c) Debriefing:

When the teacher is available, have a short conversation with them.

- Remind them of the purpose: this is to understand the lesson from their point of view, not to provide feedback or evaluate them
- Remind the teacher that you will be using an audio-recorder:

“As we’d mentioned before, I am going to be recording our conversation so we can use it later. The first thing I will do is state some details including your name. This is so we can identify the audio segment later, but in all transcripts that name will be replaced by a code. The audio file will be kept in a password protected drive and not shared outside of the team.”

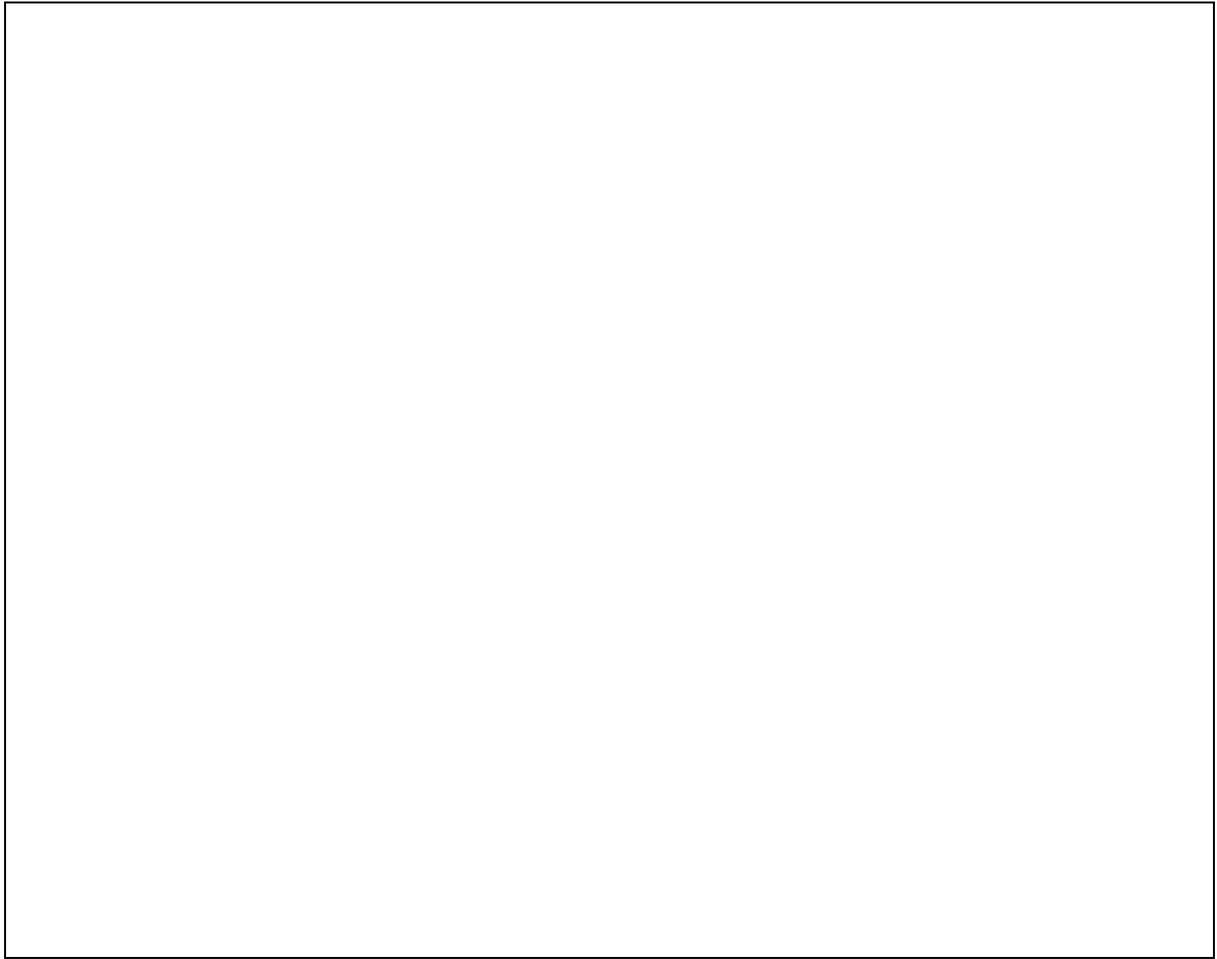
- Remember to start the audio recorder and tell the teacher when you do.
- State your name, the date, name of teacher and the lesson to be debriefed
- Debriefing questions/ points to cover – ask the teacher:
 - o to tell you about the lesson, e.g. “To get us started, could you describe the lesson to me, as though you were telling someone who wasn’t there what the lesson was about and what the class did”
 - o about what their intention was for the lesson, e.g. if there was no stated ‘learning objective’, ask, “What did you want the children to learn or practice in this lesson?” and then ask, “In addition to your learning objective, was there anything in particular you wanted to achieve or try out with this lesson?”

- whether anything was different from their original plan, and why, e.g. “How did the lesson compare to what you had planned?”
 - whether anything surprised them, e.g. “Did anything happen in the lesson that you thought was surprising or unexpected, either good or bad?”
 - about challenges and affordances, e.g. “What advice would you give to another teacher who wanted to do a similar lesson?” – if they do not give examples of what worked and didn’t work, ask, “If you were to do this lesson again, with the similar children or with a different group of children, what would you change and what would you keep the same?”
- Tell the teacher you are going to stop recording and press stop. What the teacher now says is off the record.
 - Thank the teacher for their time and for letting them see their classroom.

Appendix F2 - Running Record Template

Note: Section 3 categories were not always used and not all categories were relevant in any given lesson.

Section 1: about the lesson		
Teacher:	School:	
Observer:	Date:	
Start time:	End time:	Subject (if any):
Overview of activities:		
Section 2: running record		



Section 3: categories

Use elaborative utterances to challenge students to think more deeply about their response

Follow up children's answers with questions that vary in specificity and difficulty depending on the amount of support students require

Maintain contact with parents regarding their child

Help children make appropriate choices by monitoring the efficacy of their choices

Stand back to observe, listen and notice children's engagement and learning

Have a closure routine to help children evaluate their work and plan what happens next

Involve children in evaluating successful examples, strategies and each other's work

Plan open-ended tasks

Plan to provide strategies for children to independently solve problems they may encounter

Allow children to use a preferred method for doing a task

Ask open-ended questions

Allow children to choose how to present their work

Provide activities that are 'low threshold, high ceiling'

Allow children to choose the topic for their work (within the context of the learning e.g more or less novel or complex topics)

Allowing divergent opinions

Organise the classroom so that tasks/games are freely accessible. Children can choose whichever task they want to do, as long as they put it back when they are done.

Allow children to move around the classroom as they wish as long as they don't disturb the learning of others

Allow children to sit where they like during activities, as long as they don't disturb the class' learning

Allow children to move to a location free of distractions, where there is more space or by moving closer so they can hear/see more easily

Have a system whereby children can choose who to work with

Have a system whereby students share strategies before asking the teacher for support

Involve students in conversations about which materials to use for different purposes

Set up the classroom so that children know where resources are, how to access them and how to put them back when they are done

Plan long, uninterrupted periods of time in which children can set their own goals and control the pace and the amount of time they spend on each task

Provide children with support if they are at risk of not completing their goals, or discuss when the task should be completed

Encourage and support children to maintain focus even when faced with distractions

Provide children with a rich context, hook or stimulus where their questions might emerge - and use these to plan

Allow children to choose the topic for their work (e.g. more or less novel or complex topics)

Involve children in deciding the purpose and goals for their learning

Respond to children's interests as they emerge by being flexible in planning lessons and by drawing on these interests as learning opportunities

Recognise and verbalise children's thoughts and feelings

Actively respond to children's cues of distress

Model and foster empathic behaviour for and among children

Encourage children by demonstrating your interest in their work

Use informational rather than directive language

Anticipate and adjust the level of support depending on your understanding of the child

Help children make appropriate choices by helping them think metacognitively

Appendix G - Coding Process

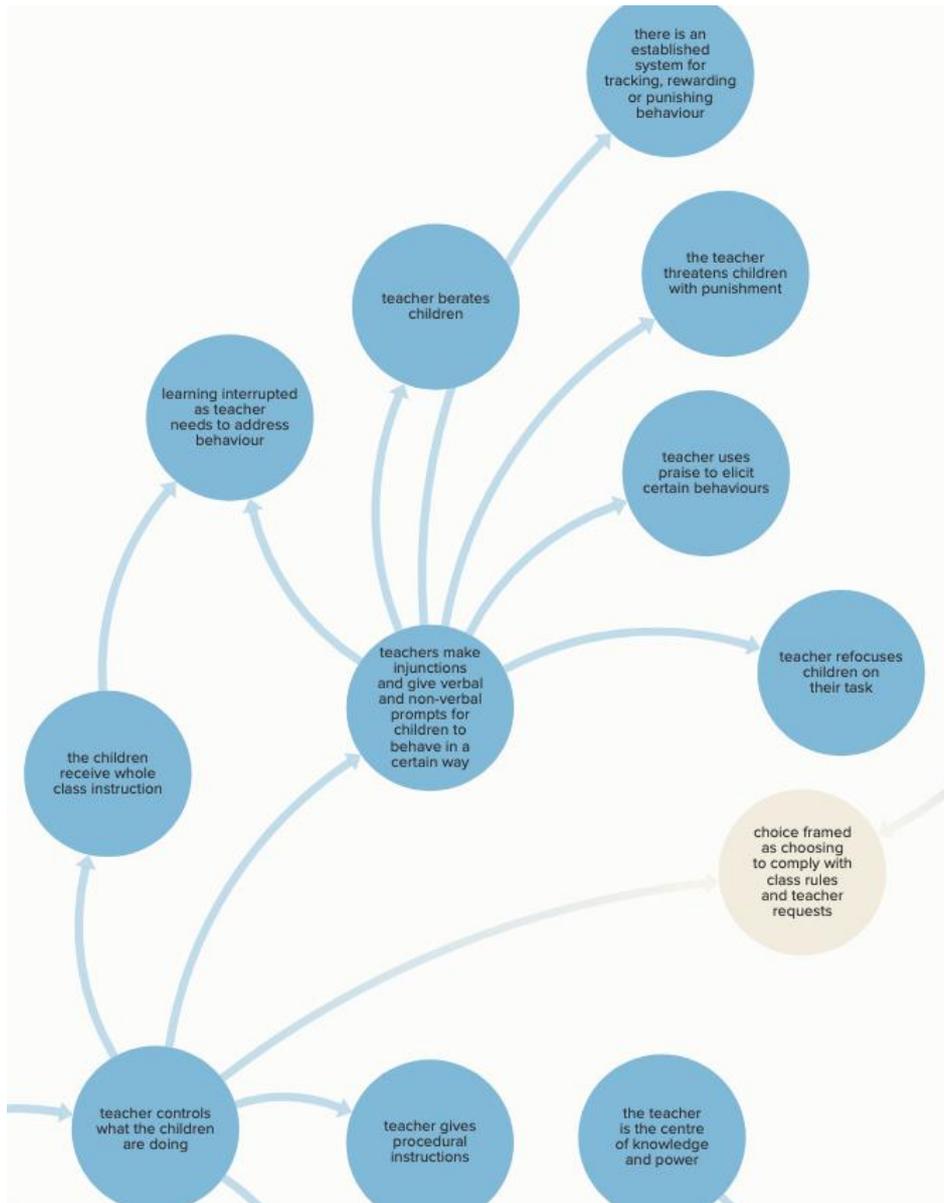
As described in the main thesis, I followed a reflexive thematic analysis method (Braun & Clarke, 2006; 2019; 2021a, 2021b), with some variation away from the prescribed steps. Here I provide more details regarding my coding process.

First, I familiarised myself with the data. This was done at the same time as checking transcriptions against audio records. Running records were read and checked again for typos. As I had carried out data collection for five of the teachers, I was already very familiar with these data.

I then carried out inductive (or open) coding. Codes or descriptions were attributed to different passages, sometimes single sentences and sometimes large paragraphs. Many sections were coded more than once. This initial stage of coding was carried out in NVivo, and followed a chronological progression, i.e. first the pre-intervention running records and interviews were coded, then the mid-point data, and finally the post-intervention data and workshop transcripts. If existing codes captured a paragraph well, the same code was used again. For example, the code for 'fixed timetables' was used 21 times. Other codes were only used once. As many codes as were needed to describe the data were generated (sometimes with some highly similar ones which were later collapsed into a single code). Some codes were clustered into higher category codes and codes regularly reviewed for this. Over 1200 codes were generated in this way.

In Braun and Clarke's (2006) initial guidance, step 3 involves searching for themes and for this they recommend using maps. However, in this study, the themes emerged from the data relatively late in the process of mapping. Instead, codes were mapped and clustered in an online freeware, Kumu, to help see patterns in the data. This allowed me to establish connections and similarities between codes, whilst adding explanations and extracts to the map. Codes were colour-coded based on different areas of interest (such as teacher control or constraints), but these did not necessarily translate into the final themes. Instead, they corresponded more to what Braun & Clarke (2006) call domains. However, the process of mapping the codes meant that similar mechanisms, or by contrast contradicting mechanisms, could be seen side-by-side. It allowed what grounded theorists have called 'axial coding', which is where analytical linkages are made between codes of data.

Figure H1 *Early Map of Codes at Stage 3, February 2020*



Alongside the coding and the mapping, I made notes of ideas for these linkages and wrote up some ‘key incidents’ which reflected what felt like important moments or important indicators of the teacher’s practice. These contributed to my thinking of what the key ideas were in the data.

This paved the way for the next step: refining and naming themes, including the need to identify what is of interest about the themes and why. However, as Braun & Clarke argue, themes are not ‘buckets’ or domain summaries (like benefits and constraints of the projects) but rather analytical units. In addition, the themes are not merely descriptive but explanatory. As I wrote the analysis, I not only drew on the data within each code, but also looked within

the dataset for episodes that further supported or which conflicted with the explanation. This also allowed me to see each extract in context to ensure it was not taken out of context.

Appendix G References

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

Braun, V., & Clarke, V. (2019). Reflecting on reflexive thematic analysis. *Qualitative Research in Sport, Exercise and Health*, 11(4), 589–597.
<https://doi.org/10.1080/2159676X.2019.1628806>

Braun, V., & Clarke, V. (2021a). Can I use TA? Should I use TA? Should I not use TA? Comparing reflexive thematic analysis and other pattern-based qualitative analytic approaches. *Couns. Psychother. Res.* 21, 37–47. <https://doi.org/10.1002/capr.12360>

Braun, V., & Clarke, V. (2021b). Conceptual and design thinking for thematic analysis. *Qualitative Psychology*. Advance Online Publication.
<https://doi.org/10.1037/qup0000196>

Appendix H - Validity in Interpretive Research

An important issue (and criticism) of interpretive research is its validity. In other words, if qualitative research is based on the researcher's insights and constructions, how can we ensure qualitative research is more than mere anecdotes or the researcher's opinion? Because of the roots of the project in more positivist-oriented research, it seems important to argue in favour of the validity of such an interpretivist endeavor.

However, before delving further into this question, it is important to note that many qualitative researchers dislike the use of the term 'validity' in interpretive qualitative research (e.g. Denzin & Lincoln, 2005) for its positivist connotation "of a singular, knowable reality" (Maxwell, 2012a, p127). Instead, many argue that terms such as trustworthy, grounded or credible should be used instead (see Whitemore et al., 2001 for a review). However, there is little consensus as to what the equivalent in qualitative research ought to be (Whitemore et al., 2001), and realists such as Maxwell (2012a) argue that the term validity can be used in qualitative research but entails differences compared to quantitative research – that it is not about some concept of a 'correct, objective reality' against which accounts can be compared, but about the relationship between an account and its object. Therefore, following Maxwell, I use the term validity to mean the extent to which conclusions satisfactorily relate to the observations made.

This involves a slightly different approach to ascertaining validity. Qualitative researchers have long discussed what might be a guarantee of quality research as understood by the community of researchers. Different strategies and criteria have been proposed, such as the use of thick description, reflexivity, triangulation and informant feedback (e.g. Schwartz-Shea, 2015). However, whilst methods are important, as Maxwell argues these procedural approaches to validity cannot on their own ensure quality in qualitative research (Maxwell, 2012a, p130). This is because these standards pertain to the methods and not to the conclusions researchers have reached (ibid). In particular, he argues validity is broadly about the plausibility of different explanations in light of existing and potential evidence. Therefore, it involves engaging not only with the conclusions, but with potential alternative explanations as well as the relationship between the conclusions and the evidence itself – i.e. questioning whether the conclusions are warranted by the data. This involves identifying plausible alternatives (or rival hypotheses, Yin, 2009), which I do for example in section 5.7 on alternative theories for explaining the tension in teachers' practice between teacher control

and opportunities for children's autonomy. This involves drawing on abduction – the appeal to the best possible explanation – as a valid logical tool (Schurz, 2008).

Examining validity also involves considering the ways in which the conclusions may be wrong, which I consider throughout, for example when discussing how teacher control may or may not have resulted in feelings of being controlled in children, or when noting that certain observations regarding curriculum pressure were not equally felt by all teachers. I also specifically focus on this in section 6.3 on limitations.

A further issue often cited when the validity of qualitative research in general is questioned is the lack of 'rigor' and 'objectivity' of such research. Qualitative researchers dismiss such criticisms because they stem from a positivist understanding of knowledge claims and the ways in which researchers arrive at such knowledge. For example, Yanow (2015) shows that the concept of rigor and its implications for research – that protocols are adhered to in an unyielding and inflexible manner – cannot apply to qualitative research because most qualitative research involves replies and acts (e.g. in an interview) that cannot be scripted in advance. However, together with other researchers she argues that this lack of rigor as 'stiffness' is not a sign that such research is not systematic. She also argues that rigor in the broader sense of logical cogent arguments does apply to qualitative research:

Here is analytic rigor: the crafting of a sound argument, in which observations build upon observations, sentences upon sentences, paragraphs and sections upon themselves, until the logic of the whole compels reason to say, Ah, yes, this makes sense as an explanation! (Yanow, 2015, p102)

Therefore *rigorous* qualitative research provides arguments that are well-evidenced and compelling. Here again, this requires relying on abduction.

The second charge, that of a lack of objectivity, is easy to accept – qualitative research always involves subjectivity – but it is the premise that research ought to be objective that qualitative researchers reject. There are a number of layers to the issue. The first is that no research is every entirely objective. For the interpretivist researcher, understanding is always produced by people, and therefore 'cognitive externality' is impossible (Yanow, 2015).

Another layer is that objectivity is not only impossible, it is undesirable in qualitative research, because it is by being human that researchers can make sense of human meaning and action. As Yanow writes: "it is our humanity, first and foremost, that enables empathic

recognition of human reaction to human experience” (Yanow, 2015, p. 105). She argues that, *contra* criticisms of interpretive research, not using human judgement in research may in fact result in research that is seriously flawed. Interestingly, it is this same observation that underpins the Leuven Involvement Scale I discuss in the second (quantitative) part of this thesis.

Finally, the ways in which the terms objectivity and subjectivity are used (largely by those critical of qualitative research) carry with them connotations: that subjective necessarily means biased or inaccurate, and therefore cannot be relied upon (Yanow, 2015).

In addition to the issues of rigor and subjectivity, I wish to address two further criticisms that are sometimes made of interpretive research: that qualitative research cannot make claims about causality, and that it cannot make claims about the generalisability of its conclusions.

With regards to the first criticism, Maxwell (2012b) in particular has argued at length that in fact qualitative research is well-placed to investigate causal relationships by attending to the process and mechanisms of causal relationships. Both realists (e.g. Pawson & Tilley, 1997) and pragmatists (e.g. Biesta, 2010) have criticised the positivist ‘regularity theory of causation’, attributed to Hume and which is the cornerstone of causal quantitative research: namely an account of causation that holds that relationships between two variables are ‘law-like’ in that a change in one will result in a change in the other. Above I have shown how such accounts, for example in RCTs, result in problematic evaluations of the quality of programmes. By contrast, realist or ‘process’ accounts of causation look for generative mechanisms, and according to Maxwell (2012b) have three important characteristics. First, they focus on process, and therefore imply that in some cases causation can be directly observed, as well as studied in single cases. Secondly, they emphasise the importance of context for understanding phenomena. And thirdly, they extend causal power to non-physical objects such as beliefs, values and meanings. Causation and mechanisms form an important part of the theory presented in Chapter 5.

Secondly, qualitative researchers have criticised the notion that it is not possible to generalise from qualitative research (e.g. Maxwell, 2021; Roald et al., 2021). Braun and Clarke, for example, argue that if qualitative research is to be of any use, we need to be able to generalise conclusions beyond the individual events or participants described in the study (Braun & Clarke, 2021). This should not be taken to mean that the theory can be taken ‘as is’

and applied indiscriminately in any context, but rather that conclusions “transcend individual idiosyncrasies” (Yanow, 2015, p107).

Appendix H References

- Biesta, G. (2010). Pragmatism and the Philosophical Foundations of Mixed Methods Research. In A. Tashakkori & C. Teddlie, *SAGE Handbook of Mixed Methods in Social & Behavioral Research* (pp. 95–118). SAGE Publications, Inc.
<https://doi.org/10.4135/9781506335193.n4>
- Braun, V., & Clarke, V. (2021). Conceptual and design thinking for thematic analysis. *Qualitative Psychology*, Advance Online Publication.
<https://doi.org/10.1037/qup0000196>
- Denzin, N. K., & Lincoln, Y. S. (2005). *The Sage handbook of qualitative research* (3rd ed.). Thousand Oaks, Calif. Maxwell, J. A. (2012a). *A realist approach for qualitative research*. SAGE, Los Angeles, Calif.
- Maxwell, J. A. (2012a). *A realist approach for qualitative research / Joseph A. Maxwell*. SAGE.
- Maxwell, J. A. (2012b). The Importance of Qualitative Research for Causal Explanation in Education. *Qualitative Inquiry*, 18(8), 655–661.
<https://doi.org/10.1177/1077800412452856>
- Maxwell, J. A. (2021). Why qualitative methods are necessary for generalization. *Qualitative Psychology*, 8(1), 111–118. <https://doi.org/10.1037/qup0000173>
- Pawson, T. & Tilley, N. (1997). *Realistic Evaluation*. SAGE Publications Ltd.
- Roald, T., K ppe, S., Hansen, J., Bechmann Jensen, T., & Levin, K. (2021). Why Do We Always Generalize in Qualitative Research. *Qualitative Psychology*, 8(1), 69–81.
<https://doi.org/10.1037/qup0000138>

Schurz, G. (2008). Patterns of abduction. *Synthese*, 164(2), 201–234.

<https://doi.org/10.1007/s11229-007-9223-4>

Schwartz-Shea, P. (2015). Judging Quality Evaluative Criteria and Epistemic Communities.

In D. Yanow & P. Schwartz-Shea (Eds.), *Interpretation and method: Empirical research methods and the interpretive turn* (2nd ed., pp. 120–146). Routledge.

<https://ebookcentral.proquest.com/lib/cam/detail.action?docID=302458>

Whittemore, R., Chase, S. K., & Mandle, C. L. (2001). Validity in Qualitative Research.

Qualitative Health Research, 11(4), 522–537.

Yanow, D. (2015). Neither rigorous nor objective? Interrogating criteria for knowledge

claims in interpretive science. In D. Yanow & P. Schwartz-Shea (Eds.), *Interpretation and method: Empirical research methods and the interpretive turn* (2nd ed., pp. 97–119). Routledge.

<https://ebookcentral.proquest.com/lib/cam/detail.action?docID=302458>

Yin, R. K. (2009). *Case Study Research: Design and Methods*. Volume 5 (4th Edition). Sage.

Appendix I - Goals and Action Points Teachers Set Themselves During the Stepping Stones Programme

Teacher	Goal at the end of Workshop 1	Action points at the end of Workshop 3
Danielle	Continuous Provision - focus on interactions and flexibility	<ul style="list-style-type: none"> - Offer a range of materials to choose from to plan/make a minibeast habitat - Pair children with nursery children
Vicky	Flexible Planning - be guided by children's interests and plan on a week-by-week basis	<ul style="list-style-type: none"> - Flexible planning - More open-ended questions
Anne	Flexible Planning	Work on achieving a good balance of choice and structure
Lisa	Elaboration - think more deeply about learning	<ul style="list-style-type: none"> - Elaboration: involve children in conversations about which materials to use for different purposes - Building on children's interests in the class story
Beth	Elaboration using her buddy system	<ul style="list-style-type: none"> - Stand back and observe more - More children's self-evaluation of their learning
Julie	Flexible planning within existing continuous provision	<ul style="list-style-type: none"> - Provide self-reflection and evaluation opportunities for children - Stand back to observe children
Claire	Flexible planning within continuous provision	<ul style="list-style-type: none"> - Allow children to choose how to present their work

		<ul style="list-style-type: none"> - Stand back to observe and notice children's engagement and learning
Stephanie	<p>Flexible Planning - carry out a series of lessons based on children's interests</p>	<ul style="list-style-type: none"> - Create a display board for children to choose pieces of work to celebrate - Make time at the end of lessons to celebrate - Teach children how to use areas by giving a compulsory activity to do at some point in the week
Helen	<p>Flexible Planning - trial a flexible plan for 2 weeks</p>	<ul style="list-style-type: none"> - Have a closure routine to involve children in choosing their next steps - Create a display wall to celebrate each child's personal work - Help children make appropriate choices in learning activities and introduce daily challenge

Appendix J - Instruments Reviewed When Selecting a Measure of Inner Motivation

(Part II)

Authors	Instrument name	Measure of	Type of instrument	Mode of administration	Age group
Baker & Scher, 2002	Motivation for Reading Questionnaire (MRQ)	Reading motivation	Survey - Likert-type scale	self-report questionnaire	6
Baker & Scher, 2002	Inventory of children's home reading activity	Reading activity	Survey - Likert-type scale	parent report questionnaire	6
Baroody & Diamond, 2013	Snapshot observation (modified from Richtie et al. 2002)	Literacy interest	Observer ratings - Likert scale	Observation	4 to 5
Baroody & Diamond, 2013	Children's Interest measure	Literacy interest	Survey - Likert-type scale	Interview with pictorial scales	4 to 5
Baroody & Diamond, 2013	N/A	Literacy interest	Survey - Likert-type scale	Parent report questionnaire	4 to 5
Baroody & Diamond, 2013	N/A	Literacy interest	Survey - Likert-type scale	Teacher report questionnaire	4 to 5
Baruch, Spektor-Levy, & Mashal, 2016	Scientific Demonstration with Documented Response Questionnaire (SDDRQ).	Scientific curiosity	Task with open-ended questions, prompted sentence completion and observation using Likert scale	Small groups	4 to 7
Bonawitz et al., 2011	N/A	Curiosity/ desire to explore	Task videotaped and coded	One-to-one	4 to 6
Daniels, Kalkman & McCombs, 2001	N/A	Perception of teacher, learning self-competence	Structured interviews with pictorial scales	One-to-one interviews	5 to 8
Doctoroff, Fisher, Burrows & Edman, 2016	N/A	Maths interest	Task, videotaped and coded	One-to-one	3 to 5.5
Doctoroff et al., 2016	Level of Interest Survey	Maths interest	Survey - Likert-type scale	Teacher report questionnaire	3 to 5.5
Downer, Booren, Lima, Luckner & Pianta, 2010	Individualized Classroom Assessment Scoring System (InCLASS)	Engagement and interactions with others	Observer ratings	Observation	3 to 5
Frijters et al. 2000	Interest in Literacy Task	Individual interest	Survey - Likert-type scale	Interview with visual scales	5 to 6
Gottfried, 1985	Children's Academic Intrinsic Motivation Inventory	Intrinsic motivation	Survey - Likert-type scale	self-report questionnaire	9+

Howard, 2002	Activity Story Apperception Procedure	Perception of tasks as learning or play	Sorting task	One-to-one	3 to 6
Hume, Lonigan & McQueen, 2015	Child Activities Preference Checklist	Literacy activity	Survey - Likert-type scale	parent report questionnaire	5 to 6
Hume, Allan & Lonigan, 2016	Literacy interest	Literacy interest	Survey - Likert-type scale	Parent report questionnaire	3 to 5.5
Jirout & Klahr, 2012	N/A	Information Gap Curiosity trait - preference for uncertainty	Task	One-to-one	4 to 5
Laevers, 1993	Leuven Involvement Scale for Young Children (LIS-YC) / Child Involvement Scale	Engagement/ involvement	Observer ratings - Likert scale	Observation	3 to 5
McDermott, Leigh & Perry 2002	Preschool Learning Behavior Scale (PLBS)	School readiness	Survey - Likert-type scale	Adult report questionnaire	3 to 5.5
McKenna et al., 1995	Elementary Reading Attitude Survey	Attitude	Survey - Likert-type scale	self-report questionnaire	6 to 12
Measelle, Ablow, Cowan, & Cowan, 1998	Berkeley Puppet Interview	Self-perception (some elements of achievement motivation)	Survey - Likert-type scale	Puppet interview	from 3
Perry & VandeKamp, 2000	N/A	Mastery or helpless motivational orientation	Survey - Likert-type scale	Puppet interview around a task	6 to 9
Piotrowski, Litman, & Valkenburg, 2014	I/D-Young Children	Curiosity trait (sensation seeking, shyness, attention)	Survey - Likert-type scale	Parent report questionnaire	3 to 8
Ponitz, Rim-Kaufman, Grimm and Curby, 2009	Observed off-task time	Engagement	Observer ratings - videotaped and times recorded	Observation and video	3 to 5
Powell et al. 2008	N/A	Engagement	Observer ratings - Likert scale	Observation	4
Stipek, Feiler, Daniels & Milburn, 1995	Enjoyment of school and school-like activities	Enjoyment (1 item) and relationship with the teacher	Survey - Likert-type scale	Interview with pictorial scales	3 to 7
Valeski & Stipek, 2001	Children's Feelings at School	Enjoyment, confidence and relationship with the teacher	Survey - Likert-type scale	Interview with pictorial scales	4 to 6
Walsh et al., 2006	Quality Learning Instrument	Classrooms, but includes observation of children's engagement and independence	Observer ratings - Likert scale	Observation	4 to 5

Appendix J References

- Baker, L., & Scher, D. (2002). Beginning Readers' Motivation for Reading in Relation TO Parental Beliefs and Home Reading Experiences. *Reading Psychology*, 23(4), 239–269. <https://doi.org/10.1080/713775283>
- Baroody, A. E., & Diamond, K. E. (2013). Measures of preschool children's interest and engagement in literacy activities: Examining gender differences and construct dimensions. *Early Childhood Research Quarterly*, 28(2), 291–301. <https://doi.org/10.1016/j.ecresq.2012.07.002>
- Baruch, Y. K., Spektor-Levy, O., & Mashal, N. (2016). Pre-Schoolers' Verbal and Behavioral Responses as Indicators of Attitudes and Scientific Curiosity. *International Journal of Science and Mathematics Education*, 14(1), 125–148.
- Bonawitz, E., Shafto, P., Gweon, H., Goodman, N. D., Spelke, E., & Schulz, L. (2011). The double-edged sword of pedagogy: Instruction limits spontaneous exploration and discovery. *Cognition*, 120(3), 322–330. <https://doi.org/10.1016/j.cognition.2010.10.001>
- Daniels, D. H., Kalkman, D. L., & McCombs, B. L. (2001). Young Children's Perspectives on Learning and Teacher Practices in Different Classroom Contexts: Implications for Motivation. *Early Education and Development*, 12(2), 253–273. https://doi.org/10.1207/s15566935eed1202_6
- Doctoroff, G. L., Fisher, P. H., Burrows, B. M., & Edman, M. T. (2016). Preschool Children's Interest, Social–Emotional Skills, and Emergent Mathematics Skills. *Psychology in the Schools*, 53(4), 390–403. <https://doi.org/10.1002/pits.21912>
- Downer, J. T., Booren, L. M., Lima, O. K., Luckner, A. E., & Pianta, R. C. (2010). The Individualized Classroom Assessment Scoring System (inCLASS): Preliminary

Reliability and Validity of a System for Observing Preschoolers' Competence in Classroom Interactions. *Early Childhood Research Quarterly*, 25(1), 1–16.

<https://doi.org/10.1016/j.ecresq.2009.08.004>

Frijters, J. C., Barron, R. W., & Brunello, M. (2000). Direct and mediated influences of home literacy and literacy interest on prereaders' oral vocabulary and early written language skill. *Journal of Educational Psychology*, 92(3), 466–477.

<https://doi.org/10.1037/0022-0663.92.3.466>

Gottfried, A. E. (1985). Academic intrinsic motivation in elementary and junior high school students. *Journal of Educational Psychology*, 77(6), 631.

Howard, J. (2002). Eliciting Young Children's Perceptions of Play, Work and Learning Using the Activity Apperception Story Procedure. *Early Child Development and Care*, 172(5), 489–502. <https://doi.org/10.1080/03004430214548>

Hume, L. E., Allan, D. M., & Lonigan, C. J. (2016). Links between preschoolers' literacy interest, inattention, and emergent literacy skills. *Learning and Individual Differences*, 47, 88–95. <https://doi.org/10.1016/j.lindif.2015.12.006>

Jirout, J., & Klahr, D. (2012). Children's scientific curiosity: In search of an operational definition of an elusive concept. *Developmental Review*, 32(2), 125–160.

<https://doi.org/10.1016/j.dr.2012.04.002>

Laevers, F. (1993). Deep Level Learning: An Exemplary Application on the Area of Physical Knowledge. *European Early Childhood Education Research Journal*, 1(1), 53–68.

McDermott, P. A., Leigh, N. M., & Perry, M. A. (2002). Development and validation of the preschool learning behaviors scale. *Psychology in the Schools*, 39(4), 353–365.

<https://doi.org/10.1002/pits.10036>

- McKenna, M. C., Kear, D. J., & Ellsworth, R. A. (1995). Children's Attitudes toward Reading: A National Survey. *Reading Research Quarterly*, 30(4), 934–956. <https://doi.org/10.2307/748205>
- Measelle, J., Ablow, J., Cowan, P., & Cowan, P. (1998). Assessing young children's views of their academic, social and emotional lives: An evaluation of the self-perception scales of the Berkeley Puppet Interview. *Child Development*, 69, 1556-1576.
- Perry, N. E., & VandeKamp, K. J. O. (2000). Creating classroom contexts that support young children's development of self-regulated learning. *International Journal of Educational Research*, 33(7–8), 821–843. [https://doi.org/10.1016/S0883-0355\(00\)00052-5](https://doi.org/10.1016/S0883-0355(00)00052-5)
- Piotrowski, J. T., Litman, J. A., & Valkenburg, P. (2014). Measuring Epistemic Curiosity in Young Children: Brief Report. *Infant and Child Development*, 23(5), 542–553. <https://doi.org/10.1002/icd.1847>
- Ponitz, C. C., Rimm-Kaufman, S. E., Grimm, K. J., & Curby, T. W. (2009). Kindergarten classroom quality, behavioral engagement, and reading achievement. *School Psychology Review*, 38(1), 102.
- Powell, D. R., Burchinal, M. R., File, N., & Kontos, S. (2008). An eco-behavioral analysis of children's engagement in urban public school preschool classrooms. *Early Childhood Research Quarterly*, 23(1), 108–123. <https://doi.org/10.1016/j.ecresq.2007.04.001>
- Stipek, D., Feiler, R., Daniels, D., & Milburn, S. (1995). Effects of Different Instructional Approaches on Young Children's Achievement and Motivation. *Child Development*, 66(1), 209. <https://doi.org/10.2307/1131201>
- Valeski, T. N., & Stipek, D. J. (2001). Young Children's Feelings about School. *Child Development*, 72(4), 1198–1213. <https://doi.org/10.1111/1467-8624.00342>
- Walsh, G., Sproule, L., McGuinness, C., Trew, K., Rafferty, H., & Sheehy, N. (2006). An

appropriate curriculum for 4–5 year-old children in Northern Ireland: comparing play-based and formal approaches. *Early Years*, 26(2), 201–221.

<https://doi.org/10.1080/09575140600760003>

Appendix K - Inner Motivation in Learning Piloted Instruments

Here I briefly describe pilot work that was carried out to design additional measures of inner motivation and reasons why this work was not pursued further.

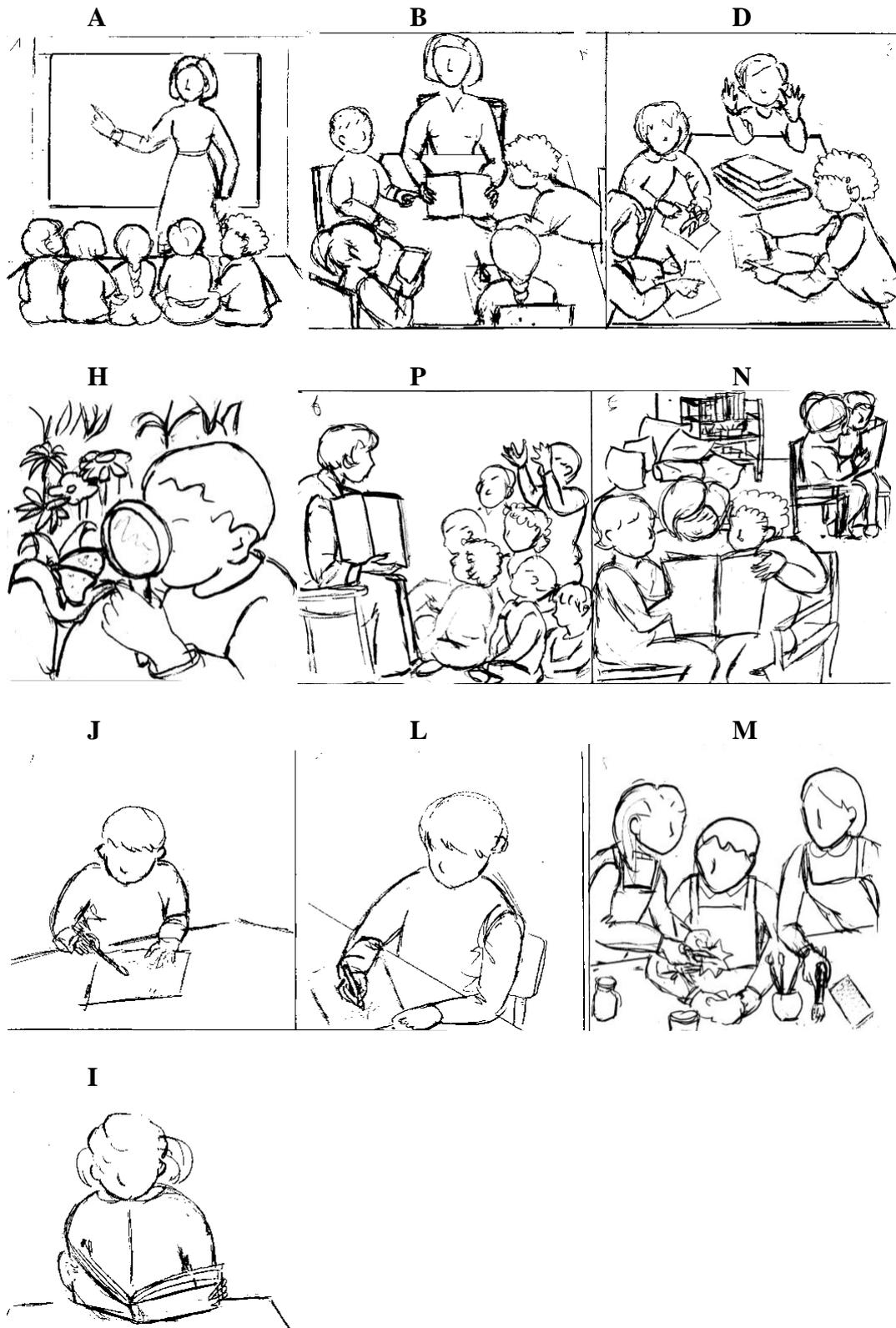
Pilot of a Forced Choice Self-Report Interview and Other Self-Report Items

In order to overcome some of the issues with self-reports outlined in Chapter 7, a new measure was designed and piloted with 8 children aged 4 to 7 in Cambridgeshire and London. The self-report measure used an interview format and targeted children's intrinsic motivation to engage in learning activities.

The measure was designed to focus on overcoming social desirability through the use of forced choice, which has been suggested as an alternative to traditional response formats because respondents are hypothesised to answer more honestly (Jackson et al., 2000). In addition, the measure used a projective technique, which asks children to interpret events using props such as pictures. Projective techniques have been used to infer children's true feelings and to avoid directly asking children to self-reflect, which they might find too challenging in certain situations (Howard, 2002; Jones, 1993; Temple & Amen, 1944).

The forced choice measure was used twice at a two-day interval to estimate reliability of the children's answers. It consisted of 10 black-and-white drawings (Figure K1) depicting different activities children might experience in a classroom (painting, class reading, group work, bug-hunting). For each drawing, the child was told, "I wonder why the children are doing ___", followed by the activity in the drawing. The children were then shown two identical bear pictures. One bear gave an intrinsic reason (e.g. "I think they are doing it because it's interesting") and the other gave an extrinsic reason (e.g. "I think they are doing it because they want to get a sticker"). The children were then asked to choose which bear they agreed with (with the order counterbalanced).

Figure K1 *Forced Choice Instrument Vignette Prompt*



Note. Numbers were erased and letters were written on the back of the pictures. Missing letters account for ambiguous vignettes that were removed following early testing prior to the pilot.

Other measures with traditional formats (open-ended questions and agree/disagree self-reports) were also trialled for comparison. An adaptive approach was used; if a measure did not provide any insights or consistent answers, alternative methods were piloted with subsequent children. This process resulted in cycles of piloting (Figure K2), including open-ended questions, traditional agree/disagree formats using the same items as the forced choice measure, and asking children to either sort the pictures or give a reason for the children’s involvement in the activities. The sorting task asked which activities the child did at school, then the ones he liked doing and finally whether they were work or play. The traditional agree/disagree formats used the identical bear faces, which either said ‘This is like me’ or ‘This is not like me’ and children were invited to choose which bear they agreed with; the order and bear were counterbalanced. Additionally, the existing Feelings at School measure (Valeski & Stipek, 2001) was used after the re-test (Day 3). Children responded to how much they agreed with statements (Figure K3) by pointing at sticks of different sizes. Although not part of the original measure, the children first practiced the response format using food items they liked and disliked. This control has been effectively used in other measures (Gottfried, 1990).

Figure K2 Summary of Piloted Measures and Process

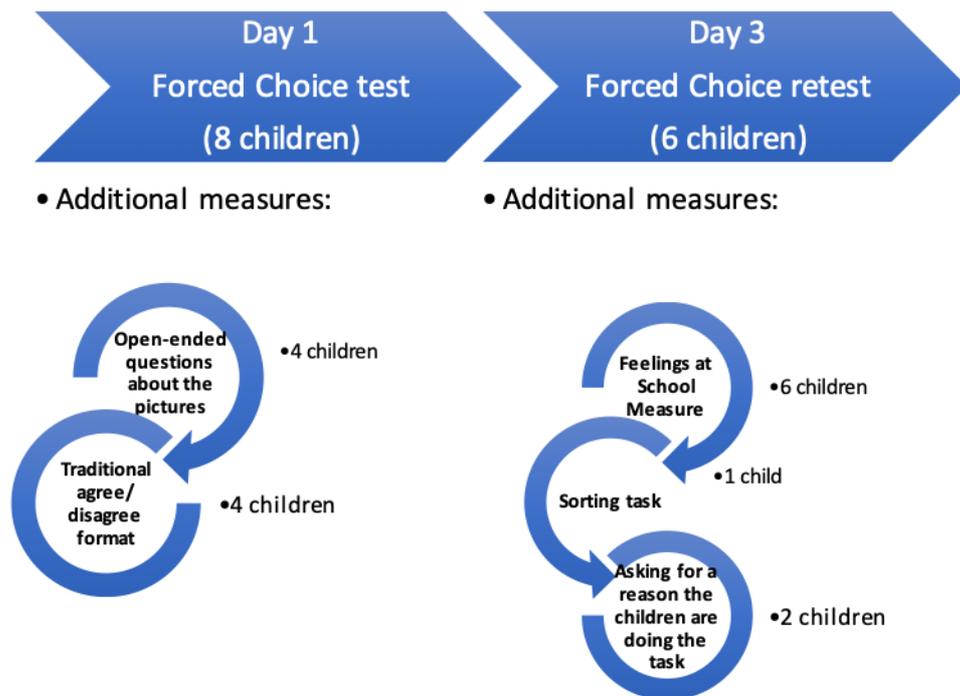


Figure K3 *The Feelings at School Measure* (Valeski & Stipek, 2001)

1. "You can use these bars to show me how much your teacher cares about you."
Point to bar 5 (largest): "If your teacher cares about you a lot, then point to this bar." Move your finger across the bars from bar 4 to bar 2: "If your teacher cares about you a little, then point to one of these bars in the middle." Point to bar 1 (smallest): "If your teacher doesn't care about you at all, then point to this bar." "Which one shows how much your teacher cares about you?"
2. "You can use these bars to show me how much you know about reading." (1 don't know much at all, 5 know a lot)
3. "You can use these bars to show me how you feel about going to school." (1 like going to school a lot, 5 don't like going to school at all)
4. "You can use these bars to show me how your teacher feels about you." (1 doesn't like you at all, 5 likes you a lot)
5. "You can use these bars to show me how fun the things you do in school are." (1 not fun at all, 5 very fun)
6. "You can use these bars to show me how good you are at numbers/math." (1 not good at all, 5 very good)
7. "You can use these bars to show me how much you know about letters." (1 don't know much at all, 5 know a lot)
8. "You can use these bars to show me how you feel when you are at school." (1 sad all the time, 5 happy all the time)
9. "You can use these bars to show me how much you know about numbers/math." (1 don't know much at all, 5 know a lot)
10. "You can use these bars to show me how good you are at reading." (1 not good at all, 5 very good)
11. "You can use these bars to show me how you feel about your teacher." (1

However, none of the measures piloted yielded responses that could confidently be considered as representative of children's motivation for school. The traditional format items gave answers clustered at the high end of the scale, with too little variability for the measures to be useful. In addition, answers to open-ended tasks were not consistent with the scores on

the scales, while the forced choice format seemed to produce a random response pattern in children. Research on children's metacognitive skills suggests that the issue lies not with children's introspective ability per se, but with the level of abstraction the forced choice measure requires, which may not be age appropriate. In addition, the pilot made apparent that the issue may not be with social desirability but with the high level of enthusiasm that children of this age have for school and learning. This means that such measures would not be sensitive enough to produce evidence of change in children's motivation in the short lifespan of an intervention.

Other Measure Considered

I considered developing a parent report of children's interest in school learning, operationalised as spontaneous involvement in the home environment with school-initiated learning. The potential of such an instrument was investigated through coffee mornings in March 2018 with Reception and Year 1 parents in two schools.

After the coffee mornings, I began to doubt how likely such a construct would be to change in children in the lifespan of an intervention, especially since children's initial interest was likely to be high. Indeed, in both schools parents reported their child had high enthusiasm for school learning, and one parent contrasted this to the level of enthusiasm of an older sibling. This is not surprising as, as previously mentioned, children have generally been found to start school with high levels of motivation. It would still be possible for such an instrument to provide more fine-grained insight into young children's interest in school learning, but considering the amount of work required to validate such an instrument it was decided that it would not be the best avenue to explore at this point in time.

Appendix K References

Gottfried, A. E. (1990). Academic intrinsic motivation in young elementary school children.

Journal of Educational Psychology, 82(3), 525.

Howard, J. (2002). Eliciting Young Children's Perceptions of Play, Work and Learning

Using the Activity Apperception Story Procedure. *Early Child Development and*

Care, 172(5), 489–502. <https://doi.org/10.1080/03004430214548>

- Jackson, D. N., Wroblewski, V. R., & Ashton, M. C. (2000). The Impact of Faking on Employment Tests: Does Forced Choice Offer a Solution? *Human Performance*, 13(4), 371–388. https://doi.org/10.1207/S15327043HUP1304_3
- Jones, R. (1993). *Perceptions of School amongst primary school pupils with and without behavioural problems*. PhD thesis, University of Wales.
- Temple, R., & Amen, E. W. (1944). A study of anxiety reactions in young children by means of a projective technique. *Genetic Psychology Monographs*, 30, 61–113.
- Valeski, T. N., & Stipek, D. J. (2001). Young Children's Feelings about School. *Child Development*, 72(4), 1198–1213. <https://doi.org/10.1111/1467-8624.00342>

Appendix L - Leuven Involvement Scale Pilot

Aims of the Pilot

The aim of the pilot was to begin testing the reliability of the Leuven Involvement Scale (see Chapter 7), and in particular its inter-rater reliability as well as what we thought of at the time as its test-retest reliability, but which we now consider as its stability over time. At the time, we were considering stability at the level of the child because that is how it had been studied elsewhere (e.g. Pascal et al, 1998; Ulich & Mayr, 2002).

The purpose was to understand whether we could use the instrument in further iterations of the Stepping Stones programme (Chapter 3), and therefore whether we could trust ratings and whether ratings would vary from one day to another – in other words, whether we could trust that ratings taken on any given day would be representative of that child's engagement in general.

When it became clear that there was an enormous amount of variation between observations, and therefore there would be issues with considering any form of child-level stability, we became interested in factors that could explain this variation. We chose activity setting as we had already collected information about this through notes on the activity context. We also initially investigated whether children had a choice of activities, but this was indistinguishable from free flow settings, so this categorisation was not pursued further.

Methods

Participants. 48 children (48% female) from 6 classes across 4 schools in England were observed in their regular classrooms. Children were in Reception (70%) and Year 1 (30%) and aged 4 ½ years to 6 ½ years (average: 5 years 4 months). All children whose parents had agreed to take part in the research participated in the study.

Procedure. Observation procedure and analytical approaches were identical to those described for the main study in Chapters 7 and 8. The main difference is that only three to four rounds of observations were carried out across one day. I coded 258 events in total, and 122 events (48%) were simultaneously rated by an additional researcher. We compared our ratings and debriefed to check our interpretations after a number of rounds; this occurred at increasingly-long time intervals over the course of the pilot, from a debrief after every round,

to a debrief every half-day; we did not alter the recorded ratings after the debriefs. For 73% of the children, I also collected data on an additional day, resulting in six observations per child instead of three.

Table L1 *Summary of the Number of Observations (Pilot)*

	1 day of observation	2 days of observation	Total
Number of schools	1	3	4
Number of classrooms	2	4	6
Number of children	12	36	48
Total number of observations	36	216	252
Number of observations per child	3	6	3-7

Table L2 *Summary of Information About the Schools and Classrooms Visited (Pilot)*

School	State/private	Percentage bands ^a of children eligible for deprivation premium (schoolwide, 2019)	Class	Year group	Number of children
A ₀	state	40-50	P1	Reception	13
B ₀	state	>50	P2	Year 1	7
			P3	Year 1	7
C ₀	state	20-30	P4	Reception	9
			P5	Reception	3
D ₀	state	20-30	P6	Reception	9

Sample Size. Sample size considerations related to the number of children and the number of observations. For inter-rater reliability, the number of observations was most important, and with over 100 observations, the sample size was sufficiently large. Stability was investigated through data collected over two days. Initially, the intention was to investigate stability at the level of the child, and therefore a sample of 48 children was sufficient.

However, because sample size had already been determined for these reasons, the sample size at higher levels was small for multilevel models. The sample size at the child level (i.e. the total number of children) was sufficient though low for fixed effects, but it made it difficult to show between-children variation compared to within-child variation. Similarly, the sample size at class level ($n=6$) was too small to show variation at the classroom level. However, these analyses were preliminary and followed up in the main study, where sample size for multilevel models was given greater consideration.

Results

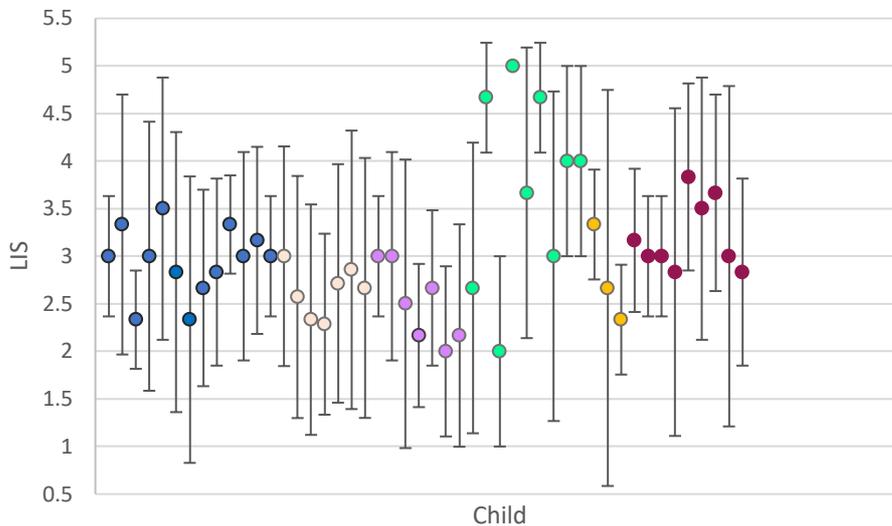
Inter-Rater Reliability. Inter-rater reliability was high ($K\text{-alpha} = .81$, $CI = 0.72 - 0.87$), which suggests that the observations made with the LIS were reliable. Inter-rater reliability was also calculated separately for the different observation sessions, and it increased from 0.68 in the first session to 0.90 in the last session. This is because in the pilot, we compared our ratings at regular intervals, and therefore our common understanding improved over time.

For comparison purposes with regards to other studies, percentage agreement within one rating of each other (the traditionally reported measure of agreement in studies using the LIS) was 93%.

Stability. Stability was poor. I used intra-class correlations (ICC) on the two-day dataset to calculate consistency between engagement ratings for each child who had been observed on two separate days. The ICC for all six observations for each child was close to 0 and non-significant ($ICC = -.023$, $F=864$, $p = .688$), which suggests that there was no consistency in the ratings for a given child when few observations are made for each child, i.e. when cluster size is small. Previous studies have used six observations split over two sessions (Bertram and Pascal, 1997), which is similar to the design of the study. These results

suggest high variability in engagement for each child from one moment to the next (Figure L1)

Figure L1 Mean LIS and Standard Deviation by Child (Pilot)



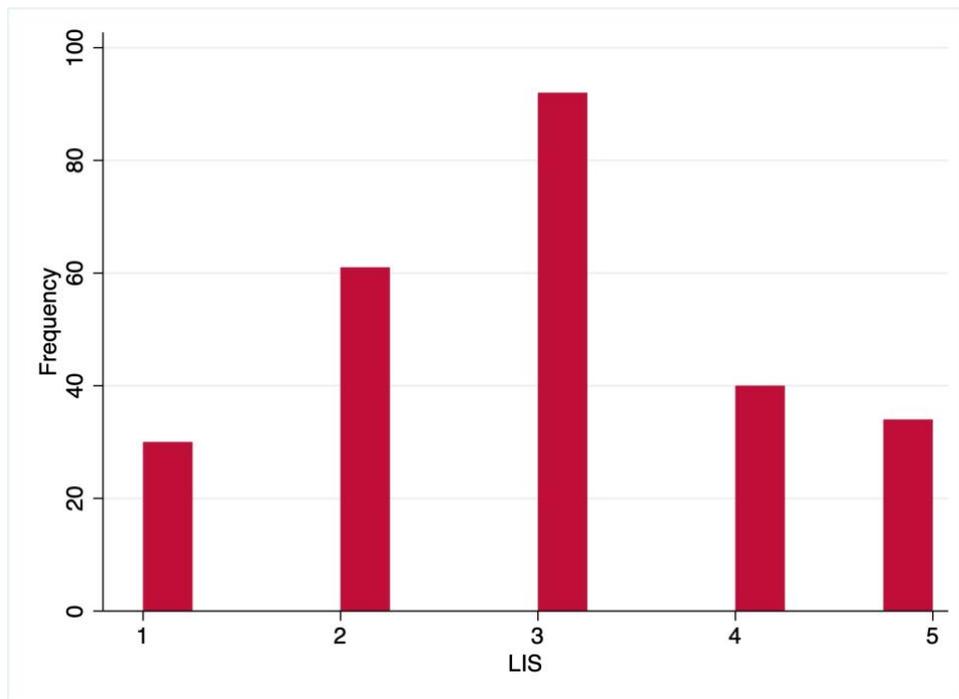
Note. Different colours represent different classrooms.

Variability. Variability was further investigated using multilevel modelling (see Chapter 8 for further details on the methods).

Descriptive Statistics. Distribution of engagement approached normality (Figure L2; skewness: 0.16; kurtosis: 2.3). The mean engagement of children was 2.95 (s.d.= 1.18) with an average engagement of 3.15 (s.d.= 1.17) in Reception and 2.57 (s.d.= 1.11) in Year 1.

Children were most often in whole class settings (46% of observations), and equally often in independent work or free flow settings (24% and 23% of observations, respectively). They were on occasion observed in other activities (6%), and never in teacher-led small group activities.

Figure L2 Frequency Distribution of LIS Scores (Pilot)



Variation Accounted for by Each Level. In the empty model, the school and the class uniquely accounted for 3% and 5.5% respectively, whilst the child uniquely accounted for no additional variation (Table L3), likely because of the small number of observations for each child. In the final model, the school accounted for 3% whilst the class and the child accounted for no variation in engagement ($ICC < 0.001$) and the likelihood ratio test suggested the model was no different from a linear model ($\chi^2 = 1.7, p = 0.63$). Again, this is likely caused by the small number of observations per child as well as the small number of classrooms. Indeed, confidence intervals were extremely large.

Table L3 Intraclass Correlation Coefficients for Each Model (Pilot)

Model	School	School: Class	School: Class: Child
empty	0.028 CI= 0.028 - 0.856	0.085 CI= 0.018 - 0.317	0.085 CI= 0.018 - 0.317
Random intercept with correction	0.030 CI= 0.002 - 0.339	0.030 CI= 0.002 - 0.339	0.030 CI= 0.002 - 0.339

Predictors of engagement. We investigated the association between engagement and activity setting, which is to say whether children were in free flow, in whole class settings,

doing independent work or other activities (see Chapter 3 and Appendix P for further details on the categories of activity settings). I also tested the effect of the control variables of age, gender and year group. None of these variables had a significant effect (Table L4) so they were dropped from further models.

Table L4 *Fixed Effects of Control Variables (Pilot)*

Study	Age	Gender (baseline: Male)	Year group (baseline: Reception)
Fixed effect	<0.01	<0.01	-0.6
Confidence interval	CI= -0.05 - 0.04	CI= -0.3 - 0.3	CI= -6.4 - 5.1
<i>p</i> -value	<i>p</i> = 0.86	<i>p</i> = 0.97	<i>p</i> = 0.31

Note. Negative values are coloured red for clarity.

Activity settings significantly predicted engagement (Table L5), as there was a significant 1.0-point increase on the rating scale when children were in free flow versus in whole class situations ($p < 0.001$). There was also a significant 0.5-point increase on the rating scale when children were in independent activities ($p = 0.004$).

For both predictor and control variables, the results reported here refer to the random intercept model with restricted maximum likelihood and Kenwood-Roger correction. The lack of between-child and between-class variability meant random slope models were not appropriate.

Table L5 *Random Intercept Model Statistics (Pilot)*

Residuals	Model parameters	Constant	Free flow	Independent	Other
1.2 CI= 0.8 - 1.8	Fixed effects	2.6 CI= 2.2 - 3.0	1.0** CI= 0.6 - 1.3	0.5** CI= 0.1 - 0.9	0.1 CI= -0.5 - 0.7
	School-level random effects	<0.1	-	-	-
	Class-level random effects	<0.001	-	-	-
	Child-level random effects	<0.001	-	-	-

Note. * $p < 0.05$, ** $p < 0.01$.

Conclusion

The pilot suggested that we could achieve good inter-rater reliability, but the lack of stability between observations and the lack of variation between children suggested much greater variation in engagement than had been assumed previous studies using the LIS, and which would make it impossible to use the LIS for child-level studies. However, because there were few observations for each child, it could have been an artefact of the small cluster size. We investigated this further in the main study by greatly increasing the number of observations per child, in order to stress-test the finding from the pilot or whether instead some pattern appeared.

Appendix M - Description of the Leuven Involvement Scale

Description of the Leuven Involvement Scale as used in this thesis and further clarifications.

The Leuven Involvement Scale

Level	Definition	Description of engagement	Examples of indicator behaviours	Clarifications
1	No engagement	Extremely low: the child shows hardly any activity , or shows <i>off-task activity</i> .	No concentration: staring, daydreaming; No goal-oriented activity, aimless actions, not producing anything; No signs of exploration or interest; Not taking anything in, no mental activity	No concentration includes signs of boredom
2	Engaged but frequently distracted	Low: the child shows some degree of activity which is often interrupted	Limited concentration, often looks away during the activity, fiddles, dreams; Is easily distracted; Action only leads to limited results	Can be scored 2 if elements of 1 and 3 are present If signs of enjoyment are present, they should be 'off-task' or not relate to goal-directed activities If actions have effects, they are not forming a coherent whole
3	Routine or passive engagement	Moderate: the child is busy the whole time but without real concentration	Routine actions ; Attention is superficial, is not absorbed in the activity; Activities are short-lived ; Limited motivation, no real dedication ; The child does not gain deep-level experiences; Does not use capabilities to their full extent, does not feel challenged; The activity does not address the child's imagination	Can be scored 3 if elements of 2 or 4 are present Child may appear sloppy or careless about the objects they're interacting with Any exploration is short-lived and not followed-through If body language is alert, it should be unprompted by an adult Activities might be repetitive or look like they are routine/ actions often repeated (e.g. chanting) and the child is doing them automatically

4	Active engagement	High: there are signs of deep absorption , but not all signals of involvement are always present to their full extent	The child is engaged in the activity without interruptions , or <i>the interruptions are brief, rare and the child refocuses themselves straight away</i> ; Most of the time there is real concentration , but during brief moments the attention is more superficial; The child feels challenged ; There is a certain degree of motivation; <i>some energy in the activities</i> The child's capabilities (and to a certain degree their imagination) is addressed in the activity	Can be scored 4 if signs of 3 and 5 are present; Signs of frustration might be present and be a sign of high challenge if the child manages them well and it leads to perseverance rather than giving up; Can show signs of exploration but not as intensely followed-up on as 5; There may be signs of enjoyment; Activity seems purposeful to a certain extent: different actions build up towards a coherent whole Signs of 4 might be present, but not any lower, and a high number of '5' statements must be correct or higher in intensity; Enjoyment may not always be visible if the child is very focused – can still rate as 5 if other signs are present at a high level; If the child shows signs of frustration or struggle with an activity, these should be minor and infrequent, and lead to renewed energy and concentration Actions are clearly purposeful
5	Intense engagement	Extremely high: during the observation of the learning, the child is continually engaged in the activity and completely absorbed in it	Is absolutely focused , concentrated without interruptions; the child is engrossed in the activity; <i>if the child is interrupted by someone else, they refocus themselves straight away</i> ; Is highly motivated, feels strongly appealed by the activity; Is not easily distracted; Is alert <i>and shows high level of energy and intensity in the activity</i> ; Shows signs that their imagination has been captured, enjoys the activity; Its mental activity and experience are intense, gains deep-level experiences; All his/ her mental capacities are in full gear , the child constantly addresses his capabilities;	

Other notes:

- Energy and intensity of actions are signs that the task matters to the child– if a task is done well but it does not seem to matter to the child, it can't be rated 5.

- Precision (trying to do well, rather than necessarily doing well) is also a sign that it matters to the child – however, children may be clumsy so it is about their attempts at trying to do well rather than how accurate their movements are. Precision can be understood as the effort and care the child puts into an activity – how much the child seems to care about the outcome.
- Exploration and imagination can also be shown through question asking (but not questions about what to do in a task or unrelated to the activity)

Some guidelines:

- The most important signals to look for are **concentration and absorption** as they are the key indicators of involvement and take precedent over other signals – if it is clear the child is deeply engrossed in an activity but other signals are absent or unclear, give priority to absorption.
- If the rater is unsure between two adjacent levels, it is better to be conservative and opt for the lower level unless one is clearly a better overall fit (even if not all statements apply).
 - Corollary: If a signal from a lower level is present, but if there are clearly more signals from a higher level, give a higher level.
- Not all signals need be present to give a rating – the rater should focus on the most salient ones.
 - Corollary: if a level asks for the presence of a signal, the rater may still give the rating if there is little opportunity to observe it or there is no obvious reason to think it is absent.
- An activity should constitute a coherent whole, even if it is made of separate actions; actions that do not relate to each other or have no purpose do not form an activity.
- Signs of enjoyment not directly related to the activity should be ignored.

Appendix N - Teacher-Researcher Memorandum of Understanding for the Leuven

Involvement Scale study

Dear _____ ,

Thank you very much for agreeing to participate in our research project. We really appreciate your willingness to support our work in your school. Here are more details about this particular piece of research, which we think will help us understand children's motivation in school. **All of the visits will be arranged in advance, at convenient times for you and your class. You won't need to change anything about your typical daily routines.**

In the autumn/spring term, a DBS-checked researcher will come to your classroom and observe randomly selected children whose parents have given permission. We will visit your classroom for **four days**. The researcher will simply watch each participating child quietly from a convenient spot in the classroom and record their level of engagement using indicators such as facial expressions and concentration, as well as some notes on what the children are doing. Please be assured that this is not an assessment of your performance as a teacher and the day of the visit should represent a typical day. The researcher will kindly ask you or a member of staff to help them identify children at the start of the day. On the fourth day, the researcher will videorecord the children as they go about their usual activities.

We will also ask you for the children's date of birth so we can calculate their exact age on the day of observation. We will also ask you to complete a questionnaire about each participating child. This should take no more than 30 minutes altogether for all the children involved.

All parents of the class will be informed in a letter about the classroom visit and child observations, which we will give to you to hand out. We will only consider children's participation if parents have returned a form giving permission for their child to take part in the study.

All data will be **kept securely** in a password-protected file with no identifier that allows a link to teachers' or children's identity, school or personal details. We replace names with codes in our files, so research information can never be linked directly to an individual's identity. Anonymised data (e.g. names replaced with codes and birth date replaced with age) may be shared with other researchers and made open for re-use. Videos will be kept in password-protected storage, which only approved members of the team will have access to. Videos will not be shared with others unless parents have given their explicit consent in the permission form. After 10 years, the video data will be deleted. You can find out more about the University guidelines on GDPR at: <https://www.information-compliance.admin.cam.ac.uk/data-protection/research-participant-data>

We are happy to share anonymised data with you for your own reflection and information. We can also provide the school with a short summary of our findings at the end of the project.

We are always available to answer any questions you may have and to discuss further details so please do not hesitate to contact me. We truly appreciate your support, and believe that together we can contribute to our understanding of children's development and learning.

Yours sincerely,

Soizic Le Courtois

PhD Student in Education

[email and phone number redacted]

Play in Education, Development and Learning (PEDAL)

University of Cambridge

Memorandum of understanding

“Stepping Stones”

I (class teacher) have read and understood that:

- A researcher will come to my class to observe children in my class for four days
- On the fourth day, the children will be videorecorded as they go about their usual activities.
- I will be asked for a list of children in my class, including their date of birth
- I will be asked to complete a questionnaire about a small number of children in my classroom
- I can withdraw from participating at any time without giving a reason

Teacher Name:	
Teacher Signature:	Date:

I (researcher) will guarantee that:

- I will consider the *Ethical Guidelines for Educational Research* the British Education Research Association (BERA) and the *Code of Ethics and Conduct* of the British Psychological Society (BPS) for conducting this research study
- I will abide by GDPR data protection legislation in the collection and storage of these research data, in particular:
 - Data will be kept confidential in password-protected files. If data are shared for re-use by other researchers, I will ensure that they also abide by GDPR legislation in their handling of the data.
 - Data files will be anonymised (e.g. names will be replaced with codes so individuals cannot be identified).
 - Videos will be kept in secure, password-protected storage.
- Research participants can withdraw without giving a reason at any time, until the data are anonymised.

Researcher Name:	
Researcher Signature:	Date:

Appendix O - Opt-In Parental Consent Forms for the Leuven Involvement Scale Study

Dear Parent or Guardian,

I am writing to you as a PhD student of the Centre for Play in Education, Development and Learning (PEDAL) at the Faculty of Education, University of Cambridge. Your child's school has kindly agreed to take part in our study about young children's motivation and engagement in classroom activities.

What is the research about?

Children's engagement is key to their learning and future success. With our research, we aim to address key questions about what motivates children to learn in school and to take part in classroom activities, and how much children's engagement varies throughout the day.

What does it involve?

*On four days, a researcher will observe the engagement in learning of randomly selected children and make notes about what activities they are doing. **On the fourth day, the children will be video recorded.** We will work closely with the class teacher to ensure that these observations do not disrupt your child's learning. We will also be asking your child's teacher for your child's date of birth so that we can identify their exact age on the day we are observing them in their classroom. We will also ask the teacher to complete a short questionnaire asking about classroom-related behaviours of each participating child.*

Does my child have to take part?

*Not at all! Participation is voluntary and **your child will only take part if you agree by returning the attached permission form.** If you change your mind, you and your child can withdraw their participation at any time before the observations. If we don't hear from you, we will assume that you do not want your child to take part in the study. We will of course also be sensitive to your child's wishes, so if they indicate that they do not want us to observe them or videorecord them, we will not do so.*

Who will see the information about my child?

Your child's information will remain strictly confidential within our team. Our project has been approved by the Faculty of Education and follows GDPR data protection guidelines. We replace names with codes in our files, so research information can never be linked directly to your child, and all data will be securely kept in a password-protected file with no identifier that allows a link to your child's identity, school or personal details. Videos will be stored in password-protected storage, which only approved members of the team will have access to. Videos will not be shared with others unless you have given your explicit consent in the permission form. After 10 years, the video data will be deleted.

Anonymised data (e.g. names replaced with codes, birth date replaced with age) may be shared for research purposes and made open for re-use. If data are shared for re-use by other researchers, we will ensure that they also abide by GDPR guidelines in their handling of the data. No personal data will ever be shared under any circumstances. You can find out more about the University guidelines on GDPR at: <https://www.information-compliance.admin.cam.ac.uk/data-protection/research-participant-data>

What are the risks and benefits of this study?

There are no known risks associated with your child's participation, given that we are simply observing them in their usual classroom activities. As a parent or guardian, you will receive no direct benefit from your child's participation in this study. However, you will help us and your child's school to understand an important element of children's learning.

We are always available to answer any questions you may have and to discuss further details so please do not hesitate to contact us. You can also contact the head teacher of your child's school or the teacher of your child's class with your questions.

We greatly appreciate your support - without it we could not do the work that we do with teachers and children!

Yours sincerely,

Soizic Le Courtois

PhD Student in Education

[email and phone redacted]

Play in Education, Development and Learning (PEDAL)

University of Cambridge

Please keep this section for your personal record.

PERMISSION FORM

*Study on children's engagement in learning
Centre for Play in Education, Development and Learning (PEDAL)*

If you wish your child to be included in the classroom observations, please complete this bottom slip and return to your child's teacher by [date].

Please tick all the boxes that apply to you. You do not have to tick all or any of the boxes.

- I agree for my child to take part in the research project on children's engagement:
 - I understand that my child will be observed by a researcher during school time.
 - I understand that my child's date of birth and some information about what they are like in school will be shared with the research team.
 - I understand my child will be **videorecorded**.
 - I understand that I can withdraw this consent at any time before data is anonymised.
 - I understand that anonymised data (no identifiable information) may be shared with other researchers.

- I agree for the videos of my child to be shared with members of the public at events or for training purposes.

_____ male female
Child's name

Signature of parent or guardian

Appendix P - Activity Setting code descriptions

Setting	Description
whole class	<p>Whole class or large group led by the teacher: Children are sat together as a class or a large group (e.g. over 10 children), and the lesson is led step-by-step by the teacher. This may at times involve independent work (e.g. on white boards) or talk but this is generally short and is always directed by the teacher. It often involves chanting and can involve watching something, doing movements or playing a game. It often involves periods of listening, including to instructions.</p>
teacher-led small group	<p>Teacher-led small group or one-to-one: The children are working with a teacher and sat either in a small group or in a one-to-one situation. Even in a small group work is unlikely to be collaborative and usually parallel with one-to-one support from the teacher, but less continuously than in a completely one-to-one situation. The task has been set or structured by the teacher, and is scaffolded by the teacher.</p>
independent	<p>Independent individual or small group work: Children are working independently from the adult, either individually or in a small group (in parallel or collaboratively) on a task set by an adult but without the adult's direct guidance. Scaffolds may exist in the structure of the task or in additional tools and resources, but is not provided by the adult. If the adult provides support, it is brief and ad hoc (e.g. a single comment) – the adult does not stay to support the child.</p>
free flow	<p>Free flow or break time: Children can freely choose between activities within a particular context (e.g. indoor play, break time, forest school). Activities include free play, but not all activities may be available and teachers may impose some restrictions. However, structured tasks set by the teacher don't count towards free flow and are instead coded as independent (if the adult isn't there) or small group (if the adult is there). Academic tasks that children choose for themselves (e.g. maths center) and that are child-led, as well as games do count towards free flow, i.e. free flow is not just free play. Arts and crafts activities are coded here unless there is an imposed outcome (e.g. create an Easter drawing) and children have not chosen to do it.</p>

Notes: Theoretically, adults may or may not be present in free flow, and may join in or even scaffold children's activities (Pyle & Daniels, 2017), but under our coding they would not initiate or structure the activity in the way they would in a 'teacher-led small group' setting (see Zosh et al., 2017); however in practice adults were hardly ever observed in free flow play. It may be possible to imagine other settings, e.g. whole class led by the students, but this never occurred.

Appendix Q - Leuven Involvement Scale Study Power Analyses

Overview

In the main study, a larger number of classrooms and children needed to be observed to establish with more certainty the impact of contextual factors and the variability between classrooms, and more observations per child were required to stress-test the finding that there was no variation in engagement between children. Therefore the minimum sample size for this must be calculated.

As different questions will require different sample sizes, these were calculated separately. It must be noted that a balance needs to be struck between the validity of claims and feasibility of data collection.

Establishing Cluster-Level Variance

Power analyses refer to effect sizes, and therefore are not appropriate for estimating the sample size needed to get a true estimate of variability between schools. Methods exist for determining the sample size needed to be representative of a given population (e.g. Krejcie and Morgan, 1970), but in this case, if we estimate the number of Reception classrooms at 34,000 (based on the 2012 number of schools in England and assuming on average schools are 2-form entries), this would entail a sample size of 380 classrooms, which is clearly unfeasible (see Table Q1). Even taking into account the child population, and limiting it to Cambridgeshire, the number of children to be observed is too high for this study and it may be difficult to recruit 30 classrooms in the area. In addition, schools who reply are unlikely to be representative of all schools, and so there is still an issue of representation. An alternative way of considering the problem is to ask how many classrooms can reasonably be recruited and observed, and judge whether this would be a large enough sample that we could be confident of the results. The Stepping Stones in Science research assistant will be available to help with approximately 10 days of data collection. It is therefore feasible that 10 Reception classrooms and 10 Year 1 classrooms could be observed, for a maximum of 20 observation days. Because schools who agree to participate are likely to give consent for both Year 1 and Reception, I would only need to recruit 5-10 schools, depending on whether they are one-form or two-form entries.

Table Q1 *Sample Size Calculations for RQ 2.1*

	Population	Sample size at the classroom level	Sample size at the child level
Classrooms in England	34,000	380	
Classrooms in each year group in Cambridgeshire	436	200	
Children in each year group in England	600,000	64	384
Children in each year group in Cambridgeshire	6,500	30	360

Note. Sample sizes assume data is collected from 12 children in each participating classroom in Reception only. Approximately double the sample size would be needed if data is also collected from Year1.

Another related question is how many children are observed in each classroom. From a practical point of view, the pilot suggests that up to 14 children can be observed three times in one school day, excluding non-classroom activities (breaktime, transitions, assemblies, PE lessons by a different teacher). In two-form entries, 6-7 children can be observed in two classrooms in one day. Observing more children per classroom would allow classroom-level variation to be more representative of the diversity of children in the classroom. However, generally it is advised to increase the higher-level sample size – one study only sampled four to eight children per classroom, but 84 classrooms (Vitiello et al., 2012). However, in this case the number of classrooms is likely to be a major constraint in terms of travel and recruitment. Therefore in order to maximise the time spent in classrooms, we may be able to either increase the number of children observed, or the number of observations per child.

Determining the Impact of Number of Observations

We could assume that a larger number of observations per child would increase between-children variability compared to within-child variability. As above, power analyses are not appropriate here. Since we're looking only for an indication that increasing the number of observations per child leads to more accurate results, there is little guidance for estimating how many such observations should be made, for how many children. Considering the level of interest here is the child, then it is important to have at least 30 children, as per 'rules of thumb' for multilevel modelling. The additional data collection required for stability of the LIS at the

classroom level will already require additional days of data collection for each child in each classroom; this presents an opportunity to increase the number of data points for each child if the number of children observed in each classroom is kept low.

Impact of Contextual Factors

In order to estimate sample size, the software MLPowSim was used, which runs simulations of power analyses for multilevel models specified by the user. The output was analysed using MLwiN. Both softwares have been built by the University of Bristol Centre for Multilevel Modelling to carry out complex multilevel models and associated power analyses. In particular, MLPowSim allows researchers to carry out power analyses for datasets that have more than two levels.

Three simulations were run with 1000 trials each with the statistics given in Table Q2 below. Across all models, the simulations were run with 6 children in each classroom, and 3 observations for each classroom (the number of observations in the pilot). Snijders (2005) notes that the cluster-level size is not very important for the power of tests focusing on the effect on a level-one variable, and that the main limiting factor is the sample size at the highest level, in this case the classroom.

The first simulation used only the statistics from the preliminary data. Explanatory variables were the three dummy variables used in the Reception-only multilevel model and corresponding probabilities were calculated using existing data. This model assumed no variance was attributable to the classroom or the child, as indicated by the variance close to 0. This explains the low sample sizes required for two of the dummy variables; because independent work had both a low effect size (-0.2 regression coefficient) and low incidence (probability of 0.2), it required a larger sample size (power of .808 for 35 classrooms).

However, because the variance close to 0 was surprising, and essentially means there is no multilevel structure, the simulation was run again with the variance from the empty model (0.062 for the classroom, none for the child) and with higher, albeit still low, variance for both classroom and child level (0.24 for the classroom, 0.084 for the child and residual variance of 0.95). These are based on a previous study looking at children's engagement with tasks in a pre-school setting (Vitiello et al., 2012) This study used a similar number of observations per child (mean = 3.7, s.d. =1.2) and of children per classroom (between 4 and 8).

Table Q2 *Sample Size Calculations for RQ 2.3 - What Accounts for Variation in Engagement*

	Model 1: ♦ classroom-level variance: null ♦ child-level variance: null ♦ observation-level variance: 1.14	Model 2: ♦ classroom-level variance: 0.06 ♦ child-level variance: null ♦ observation-level variance: 1.14	Model 3: ♦ classroom-level variance: 0.24 ♦ child-level variance: 0.084 ♦ observation-level variance: 0.95
Whole class probability 0.4; beta ₁ = -1	<10	<10	<10
Independent work probability 0.2; beta ₂ = -0.3	35	40	30-35
Other probability 0.1; beta ₃ =-0.8	<10	10-15	<10

Note. Sample size calculations are based on the number of classrooms needed to exceed a power of 0.8. The default activity for the dummy variables is free flow. The statistics for the model were derived from the Reception-only multilevel random intercepts model. The fixed intercept was beta₀=3.7.

It is debatable whether the power needed to establish effect for this specific dummy variable warrants the additional data collection. Alternatively, we could re-categorise the activity structures; this may result in a higher incidence of the activity and therefore increase power, or it could dilute the effect further and reduce power.

To see how the categorisation might affect sample size, the same three simulations were run again but this time using the information from the choice/no choice classification of activities. However, possibly because of the high incidence (0.5) and high effect size (0.8 point increase in engagement with choice), the simulations predicted a power of 1 for sample sizes of less than 10 classrooms, even when higher variability between classrooms and students was given. Running the simulation again with a lower effect size (0.5) produced very similar results.

These models assume that random slopes are not calculated. This means that it is assumed classrooms do not differ in the effect that different settings have on children – for example, whole class settings result in, on average, lower engagement equally across all classrooms. This might not be the case in situations where teachers use very different approaches. If this was the case, power might be limited by the low cluster size (Snijders, 2005), i.e. the number of observations per child, and the number of children per classroom.

In summary, the minimum number of classrooms for the multilevel analyses would be 35 in order to detect effects of independent work on engagement, but if this variable is considered less important, or if the effect becomes more prominent as we collect more data, 10 classrooms seem sufficient.

It must be noted that these simulations were run using the models fitted to the Reception-only data, and therefore this number concerns the number of Reception classrooms.

Summary

Although there are three areas of investigation, this does not mean data must be collected separately. Table Q3 describes the requirements for each question. We can see that if 20 classrooms are observed (10 in each year group), with 6-7 children in each classroom observed 3-6 times over one day, this fulfils the requirements for investigating the effect of contextual factors. It will also provide a starting point to investigate between-classroom variability. Observing these classrooms for additional days would increase the number of observations per child (i.e. cluster size) and allow us to investigate stability across different days.

Table Q3 *Summary of sample sizes needed for each area of investigation*

Area of investigation	No of classrooms (Reception/ Year 1)	No of children per class	Total number of children	No of days of observations	No of observations per child	Total number of observations
Stability	As many as possible	-	-	3	-	-
Cluster-level variance	as many as possible	unclear	-	-	as many as possible	-
number of observations per child	15 (8/7)	6	-	2	9-12	810
contextual factors	10 or 35 in Reception	6	-	1	3	-

Appendix Q References

- Krejcie, R. V., & Morgan, D. W. (1970). Determining Sample Size for Research Activities. *Educational and Psychological Measurement*, 30(3), 607–610.
<https://doi.org/10.1177/001316447003000308>
- Snijders, T. A. B. (2005). Power and Sample Size in Multilevel Linear Models. In *Encyclopedia of Statistics in Behavioral Science*. American Cancer Society.
<https://doi.org/10.1002/0470013192.bsa492>
- Vitiello, V. E., Booren, L. M., Downer, J. T., & Williford, A. P. (2012). Variation in children's classroom engagement throughout a day in preschool: Relations to classroom and child factors. *Early Childhood Research Quarterly*, 27(2), 210–220.
<https://doi.org/10.1016/j.ecresq.2011.08.005>

Appendix R - Leuven Involvement Scale Study Multilevel Model Statistics

Table R1 Full Model Detailed Statistics

model	Residuals	Model parameters	Constant	Free flow	Independent	Small group teacher-led
empty	1.2 CI= 1.0 - 1.3	Fixed effects	3.0	-	-	-
		School-level random effects	<0.1	-	-	-
		Class-level random effects	<0.1	-	-	-
		Child-level random effects	0.1 CI= 0.0 - 0.1	-	-	-
Random intercept	1.1 CI= 1.0 - 1.1	Fixed effects	2.6 CI= 2.5 - 2.8	0.7** CI= 0.6 - 0.8	0.2** CI= 0.1 - 0.4	0.3** CI= 0.1 - 0.5
		School-level random effects	<0.1	-	-	-
		Class-level random effects	<0.001	-	-	-
		Child-level random effects	0.1 CI= 0.0 - 0.1	-	-	-
child-level random slopes	1.0 CI= 0.9 - 1.1	Fixed effects	2.6 CI= 2.5 - 2.8	0.7** CI= 0.6 - 0.9	0.2 CI= 0.0 - 0.4	0.3* CI= 0.1 - 0.6
		School-level random effects	<0.1	-	-	-
		Class-level random effects	<0.001	-	-	-
		Child-level random effects	0.1 CI= 0.0 - 0.1	0.1 CI= 0.1 - 0.3	0.2 CI= 0.1 - 0.6	0.2 CI= 0.0 - 0.7
Class-level random slopes	1.0 CI= 1.0 - 1.1	Fixed effects	2.6** CI= 2.5 - 2.8	0.7** CI= 0.6 - 0.9	0.2 CI= -0.2 - 0.5	0.3 CI= -0.1 - 0.7
		School-level random effects	<0.01 CI= 0.0 - 0.2	-	-	-
		Class-level random effects	<0.01 CI <0.01	<0.1 CI= 0.0 - 0.3	0.1 CI= 0.0 - 0.7	0.2 CI= 0.0 - 0.8
		Child-level random effects	0.1 CI= 0.0 - 0.3	-	-	-

Class- & child- levels random slopes	1.0 CI= N/A	Fixed effects	2.6 CI= 2.5 - 2.8	0.7** CI= 0.6 - 0.9	0.2 CI= -0.2 - 0.5	0.3 CI= -0.1 - 0.7
		School-level random effects	<0.1 CI= N/A	-	-	-
		Class-level random effects	<0.01 CI= N/A	<0.01 CI= N/A	0.1 CI=N/A	0.2 CI=N/A
		Child-level random effects	0.1 CI=N/A	0.1 CI=N/A	0.1 CI=N/A	0.1 CI=N/A

Note. * $p < 0.05$, ** $p < 0.01$. Models were run with restricted maximum likelihood.

Table R2 Likelihood Ratio Tests for the Maximum Likelihood Models

	Empty	Random Intercepts	Random slopes child level	Random slopes class level	Random slopes child and class levels
Random intercepts	Chi ² = 146.1 $p < 0.001$	-	Chi ² = 26.44 $p < 0.001$	Chi ² = 15.7 $p = 0.001$	-
Random slopes child level					Chi ² = 5.26 $p = 0.072$
Random slope class level					Chi ² = 16.02 $p < 0.001$

Appendix S - Multilevel Models Assumption Checks

One of the benefits of multilevel models is they hold few of the assumptions of general linear models, such as homoscedasticity (homogeneity of variance) and independence of observations. In this case, the outcome variable (LIS score) was normally distributed, but this is not in fact a requirement of linear regressions; the assumption is rather that *sampling distribution* is normal, and according to the Central Limit Theorem at large sample sizes this will be the case regardless of the distribution of any given variable in the population or the sample (Lumley et al., 2002). With over 1000 observations, the Central Limit Theorem therefore applies to this dataset.

Multilevel models do make a number of assumptions about the models. Specifically, multilevel models assume: linearity in the relationship between outcome and predictor variables; homogeneity of variance and normal distribution *of residuals* at level 1; and homogeneity of variance and normal distribution *in residuals* at higher levels (Hox, 2013). As with general linear models, it is also assumed that the models have been correctly specified (Snijders & Bosker, 2012).

Visual methods were used to check assumptions of the model, as per Snijders & Bosker (2012), as assumption checking of multilevel models can be difficult without specialist software (Hox, 2013). The models did not seem to violate assumptions of multilevel modelling. Residuals at level 1 showed near-normal distribution and q-plots looked as expected; there were some small deviations from normality at the tails, but these are considered acceptable (Snijders & Bosker, 2012). Level 1 residuals were normally distributed within individual classes as well; when inspected individually for each child, there were some deviations from normality in some cases, but in general residuals were normally distributed. There were no important differences in the confidence intervals of these residuals between children and between classes, which suggests that the model was equally predictive for different children or different classes.

I also examined the distribution of random effects predictions (i.e. the BLUPs - Best Linear Unbiased Predictions for random effects), that is to say the distribution of random intercepts and random slopes at each level. The random intercepts at school level and class level were not normally distributed; this was likely caused by the low sample size at those levels. The random slopes for independent work and for teacher-led small group settings did not look like they had normal distributions (both at child level because of the model

specification). The distribution suggested that the effect of independent work did not vary for most children, but that the effect was strong (either positively or negatively) for a smaller number of children. This explains the large confidence intervals for those random slopes. Random effects were normally distributed for the child level for free flow and whole class settings (the intercept at child level). Overall this suggests that the model was not a very good fit for the class level and for individual work and teacher-led small group random effects, but was a good fit for all fixed effects, as well as for the free flow and whole class random effects at the child level. Sample size (both the number of classes, and the number of observations in independent work and teacher-led small-group work) may be the reason, and it may also be that the variables of independent work and teacher-led work were not capturing enough of what really made the difference to engagement for different children or in different situations. This is discussed in more detail in Chapter 9. Results were similar for the class-level random slope models.

One other important unmet assumption was that about data type: the outcome variable was ordinal rather than continuous. However, multilevel ordered logistic models are complex, computationally demanding, and difficult to interpret. In addition, it has been shown that the bias introduced by treating categorical data as continuous is small when there are five or more categories (Hox, 2010), as was the case here. Multilevel ordered logistic regressions were nonetheless run as a check. The direction, significance level and respective magnitude of the effect sizes of the different variables were comparable to the models presented above, though only random intercept models were able to be run as random slopes were too computationally demanding (Tables R1 and R2).

Table S1 *Fixed Effects in the Random Intercept Model (LIS Study)*

study	Free flow	Independent	Small group teacher-led	Other	Cut 1	Cut 2	Cut 3	Cut 4
Pilot study: random intercept model	1.0** CI= 1.0 - 1.4	0.5** CI= 0.1 - 0.9	-	-	- CI= -0.5 - 0.7			
Main study: Random intercept model*	1.2** CI= 1.0- 1.5	0.24* CI= 0.1 - 07	0.3* CI= 0.1- 0.6	-	-1.8 CI= 0.1- 0.6	-0.1 CI= -0.3 - 0.1	-1.2 CI= 1.0 - 1.5	-3.2 CI= 2.9 - 3.5

Note. * $p < 0.05$, ** $p < 0.01$.

Table S2 *Random Effects in the Random Intercept Model (Main LIS Study)*

Model parameters	School level	Class level	Child-level
constant	<0. CI= 0.0- 0.3	<0.01 CI= N/A	0.2 CI= 0.1 - 0.4