

Classroom behaviour management to support children's social, emotional, and behavioural development

Thesis submitted for the degree of Doctor of Philosophy in Social

Intervention

Elizabeth Nye

Candidate Number: 894383

St. Catherine's College, University of Oxford

Hilary Term 2017

Word count: approx. 85,000



Abstract

Classroom behaviour management to support children's social, emotional, and behavioural development

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

Elizabeth Nye

St. Catherine's College, University of Oxford

Hilary Term 2017

Introduction: Children's social, emotional, and behavioural difficulties are associated with reduced academic performance, stressed teacher-child relationships, and other negative academic and life outcomes. The Incredible Years Teacher Classroom Management (IY TCM) programme is one intervention developed to address problematic behaviours via training teachers to use positive and proactive management strategies.

The overall aim of this DPhil is to use the Incredible Years Teacher Classroom Management programme as a case study for applying mixed methods at the systematic review level to ascertain what is known about both the programme's effectiveness and how people experience the course, and subsequently to use the systematic review's findings as a springboard (rather than as an end goal) for more exploratory research into 'for whom' the programme might work.

Method: Study One is a mixed methods systematic review of IY TCM. It applied multilevel meta-analysis to RCT outcome data and grounded theory meta-synthesis to interview and focus group data on stakeholders' experiences of IY TCM. Quantitative and qualitative findings were cross-synthesised and mapped using an integrative grid.

Study Two moves the field forward by filling a gap in the evidence base, as identified in Study One. Semi-structured telephone interviews were conducted with special educational needs coordinators (SENCOs) across Devon, exploring the acceptability and appropriateness of expanding IY TCM to the subgroup of children with special educational needs (SEN) in mainstream schools. Data were analysed thematically and mapped onto IY TCM content.

Results: In Study One, nine studies reported across 14 papers met inclusion criteria for either quantitative or qualitative strands of this systematic review. Multilevel meta-analysis of RCTs ($n = 4$) indicated that the programme produced teacher- and child-level results in the desired directions. Clear trends across all measured outcomes favoured the intervention group over the treatment-as-usual comparison. Qualitative meta-synthesis ($n = 5$) illuminated a cyclical learning process and broader conceptualisation of teacher and child outcomes than was evident in the quantitative evidence. Notably, RCT data on teacher outcomes were limited to self-reported or observed behaviours, while teachers described other benefits from IY TCM including increased knowledge and emotional well-being. Cross-synthesis of findings from the two review strands highlighted harmony across the RCT and qualitative evidence but also a number of areas in which constructs that were prioritised by one type of research were not integrated into the other.

Study Two generated classroom management strategies from SENCOs, which aligned closely with strategies taught in IY TCM, indicating that IY TCM would be both

acceptable and applicable (if not sufficient) for use when working with children identified with SEN and behavioural difficulties in schools.

Discussion: Based on the positive effects of implementing IY TCM despite very few studies to power analyses, the programme appears to offer tangible benefits to both teachers and children. It is possible that results are underestimated due to limited types of outcomes measured and absence of experiential data from additional stakeholders (e.g., parents). Depending on current provision of special educational needs services, schools operating inclusion models are likely to find these strategies beneficial for children identified with SEN, and this subgroup should be explicitly examined in future IY TCM studies.

Acknowledgements

I wish to express my heartfelt thanks to Prof Frances Gardner for supervising me during this DPhil journey. In particular, I have valued our joint writing days for both the motivation they provided and as wonderful opportunities to work alongside one another. Thank you for investing considerable time and energy into my development as a researcher throughout the years. The insights and experiences you have shared with me will continue to guide me moving forward in my career.

During the course of the past five years, I have had the unique pleasure of experiencing the Department of Social Policy and Intervention and the Centre for Evidence-Based Intervention first as an MSc student, then as a DPhil candidate, and most recently also as a Departmental Lecturer. This progression offers me multiple reasons to extend my appreciation not only to the department as a whole but also to the many members who have most directly influenced me. Thank you to Shreya for being an incredible study partner during the MSc and a close friend in the years ever since. Many thanks go to all of my fellow DPhil candidates for being such a supportive and collaborative community of scholars. In particular, I am appreciative for Aase, Ani, Ellie, Felix, Julie, and Madeleine, as S17 would not have been the same without you. Thank you to my wonderful post-doc friends and colleagues: Franziska, G.J., Patty, and Thees. I am very grateful for all of the methodological, statistical, and professional advice you have offered, and yet it is your friendship that I treasure the most. Thank you to Jenny for fully bringing me into the departmental fold during my first year on the DPhil and for the many conversations across the hall. Finally, I would like to offer great thanks to all of my colleagues on the EBSIPE teaching team. I have really enjoyed working with you, and I have learned so much through the examples you set. I cannot imagine having started my academic career any other way.

My sincerest thanks also extend to Prof Tamsin Ford and the entire STARS Study research team in Exeter for facilitating my introduction to the schools with whom you had already cultivated close relationships. I will never forget the long days of school visits to the far corners of Devon, which were great fun and will remain one of the highlights of my DPhil research. I am extremely grateful to all of the special educational needs coordinators and

teaching assistants who took time from their very busy lives to share their experiences and views with me. I would also like to extend my thanks to the Broadbent family for warmly welcoming me into their home during my research trips to Exeter.

When I first arrived in Oxford and made my way to St. Catherine's College, I had no idea just how instrumental the Catz community would be for me. In particular, I will always think fondly of the Sunday morning welfare waffle brunches, and I wish to thank my fellow welfare officers and peer supporters for being such a genuine and caring group. A massive thank you is due to Anne, John, Samantha, and Tyson for your close friendship and our shared adventures through the years.

This DPhil journey began long before I arrived in Oxford. Grandfather and Grandmother, your love of international travel made a lasting impression on me, and I have appreciated your encouragement to follow my dreams wherever they take me. Grammy and Grandpa, your regular presence in my life has shaped me into the person I am today and is a gift for which I am eternally grateful. Mom, Dad, Will, and John, words are not sufficient to express my thanks for your love and support throughout the years. Despite the long distance and many time zones that separated us, know that I have felt your presence daily, and it encourages me in all that I do.

Mom and Dad, thank you for instilling in me a love of learning, an appreciation for hard work, a curiosity to explore different opportunities, and the confidence to seek out new challenges. You are truly an inspiration to me, and I dedicate this DPhil thesis to you.

Table of Contents

Abstract.....	i
Acknowledgements	iii
List of Figures.....	ix
List of Tables	xi
INTRODUCTION	1
Chapter 1 : Children’s social, emotional, and behavioural health difficulties – Intervening at the teacher-level in schools.....	3
Introduction	3
Defining and diagnosing social, emotional, and behavioural disorders	5
Prevalence of childhood mental health disorders	9
Special educational needs and inclusive education	10
Defining and diagnosing special educational needs	11
Prevalence of children identified with SEN and their trajectories	16
Teachers’ difficulties with classroom management	18
Incredible Years Teacher Classroom Management Programme	20
Theory of intervention	21
Mixed methods research	24
Study One	25
Study Two	30
STUDY ONE.....	35
Chapter 2 : Methods	37
Introduction	37
Research questions	37
Criteria for Considering Studies for this Review – Quantitative Strand	38

Criteria for Considering Studies for this Review – Qualitative Strand	42
Search Methods for Identification of Studies – Quantitative and Qualitative Strands ...	44
Data Collection – Quantitative and Qualitative Strands	46
Data Extraction and Analysis – Quantitative Strand	47
Data Extraction and Analysis – Qualitative Strand	56
Cross-synthesis	58
Chapter 3 : Effectiveness of a teacher classroom management programme – Results of the quantitative systematic review and multilevel meta-analysis	61
Results	61
Description of studies	61
Risk of bias in included studies	73
Effects of interventions	77
Discussion	89
Conclusion	91
Chapter 4 : Experiences of the IY TCM programme from key stakeholders – Results of a qualitative systematic review and meta-synthesis	93
Results	93
Description of studies	93
Quality appraisal for included studies	99
Experiences of key stakeholders	103
Discussion	142
Conclusion	144

Chapter 5 : Cross-synthesis of evidence from trial effectiveness and stakeholders' experiences	145
Results	145
Overview of findings	146
Conclusion	165
STUDY TWO	171
Chapter 6 : Methods	173
Need for the Current Study	173
Aims	176
Study Design	176
Sample	177
Data Collection and Management	177
Semi-structured interviews	178
Procedure	179
Consent and ethics	180
Analysis	181
Chapter 7 : Acceptability and applicability of a teacher classroom management programme for children with special educational needs in mainstream schools – Results from semi-structured qualitative telephone interviews	183
Descriptive findings	183
Overview of findings	184
Three types of problematic behaviours	185
Approach to classroom behaviour management: SENCos first seek to determine a cause	188

Approach to classroom behaviour management: SENCoS' strategies.....	190
SENCo strategies in relation to IY TCM strategies.....	207
Discussion.....	208
DISCUSSION.....	213
Chapter 8 : Discussion.....	215
Introduction.....	215
Summary of main findings.....	216
Conflicts of interest and reflexivity	223
Strengths	226
Limitations	228
Novel contribution and implications.....	233
Conclusion	235
References.....	237
Appendix A: PROSPERO Protocol for Study One.....	259
Appendix B: R Syntax for multilevel meta-analysis for Study One.....	271
Appendix C: First-authored peer-reviewed journal article on qualitative meta-synthesis associated with Study One.....	275
Appendix D: Risk of bias tables for quantitative strand of Study One.....	309
Appendix E: Quality appraisal tables for qualitative strand of Study One.....	321
Appendix F: SENCo topic guide for Study Two.....	335
Appendix G: SENCo information sheet for Study Two.....	337
Appendix H: Standard operating procedure for consent for Study Two.....	341
Appendix I: Ethics approval for Study Two.....	343

List of Figures

Figure 1.1 Illustration of the convergent parallel mixed methods design for this study	28
Figure 2.1 Research questions for this mixed methods systematic review	38
Figure 2.2 Electronic searches	45
Figure 2.3 Search strategy	46
Figure 2.4 Multilevel analysis in primary studies	48
Figure 2.5 Multilevel meta-analysis with first level implied.....	51
Figure 2.6 Excerpt of R syntax	52
Figure 3.1 PRISMA flowchart	64
Figure 3.2 Risk of bias assessment.....	73
Figure 3.3 Multilevel meta-analysis of teacher negative classroom management strategies	81
Figure 3.4 Multilevel meta-analysis of teacher positive classroom management strategies	81
Figure 3.5 Multilevel meta-analysis of child conduct problems	82
Figure 3.6 Sensitivity analysis of child conduct problems.....	83
Figure 3.7 Multilevel meta-analysis of high-risk child conduct problems.....	83
Figure 3.8 Multilevel meta-analysis of child prosocial behaviours.....	84
Figure 3.9 Sensitivity analysis of child prosocial behaviours	85
Figure 3.10 Multilevel meta-analysis of high-risk child prosocial behaviours	85
Figure 3.11 Multilevel meta-analysis of child behaviour difficulties	88
Figure 3.12 Sensitivity analysis of child behaviour difficulties	88
Figure 3.13 Multilevel meta-analysis of high-risk child behaviour difficulties.....	89
Figure 4.1 Quality appraisal of included studies	100
Figure 4.2 Result of the meta-synthesis: Cycle of learning.....	107

Figure 4.3 Result of the meta-synthesis: Teacher and classroom outcomes.....	115
Figure 4.4 Result of the meta-synthesis: Child outcomes.....	124
Figure 4.5 Result of the meta-synthesis: Critiques of the programme	130
Figure 5.1 Example of visual illustrations of cross-synthesis findings	146
Figure 5.2 Summary of qualitative findings mapped onto quantitative constructs	148
Figure 5.3 Integrative grid – quantitative constructs and qualitative studies	149
Figure 5.4 Integrative grid – qualitative constructs and quantitative studies, part one	156
Figure 5.5 Integrative grid – qualitative constructs and quantitative studies, part two	157
Figure 5.6 Summary of quantitative findings mapped onto qualitative constructs	158
Figure 5.7 Summary of quantitative findings mapped onto qualitative construct of 'critiques'	162
Figure 6.1 Subgroup of interest for questions in qualitative interviews: children identified with both SEN (Parsons & Platt, 2013) and CD (H. Green et al., 2005).....	173
Figure 7.1 Overview of findings from SENCo interviews	184
Figure 7.2 Strategies for behaviour management as described by SENCos.....	191

List of Tables

Table 3.1 Included studies	65
Table 3.2 Characteristics of included studies	69
Table 3.3 Excluded studies	72
Table 3.4 Multilevel meta-analysis results for primary and secondary outcomes	80
Table 3.5 Multilevel meta-analysis results for post hoc outcomes.....	86
Table 4.1 Included studies	93
Table 4.2 Characteristics of included studies	98
Table 4.3 Excluded studies	99
Table 4.4 Example of coding tree for teacher outcomes theme	102
Table 4.5 Example of coding tree for critiques theme	103
Table 5.1 Themes from quantitative and qualitative strands	148
Table 7.1 Example of coding tree for management strategies	185
Table 7.2 SENCo-generated strategies mapped alongside IY TCM strategies	208

INTRODUCTION

Chapter 1 : Children’s social, emotional, and behavioural health difficulties –

Intervening at the teacher-level in schools

Introduction

This dissertation focuses on how educators can support the social, emotional, and behavioural needs of children in mainstream schools. The overall aim of this DPhil is to use the Incredible Years Teacher Classroom Management programme as a case study for applying mixed methods at the systematic review level to ascertain what is known about both the programme’s effectiveness and how people experience the course, and subsequently to use the systematic review’s findings as a springboard (rather than as an end goal) for more exploratory research into ‘for whom’ the programme might work.

Children’s mental health issues, including conduct, hyperkinetic, and emotional disorders, are of international consequence. In September of 2015, the General Assembly of the United Nations adopted a resolution with its post-2015 Agenda for Sustainable Development (Assembly, 2015). Of the 17 Sustainable Development Goals, two stand out as directly pertaining to children’s health and education. Goal 3 aims to ‘ensure healthy lives and promote well-being for all at all ages’ (Assembly, 2015, p. 14). More specifically, target 3.d specifies that the international community should seek to ‘strengthen the capacity of all countries, in particular developing countries, for early warning, risk reduction and management of national and global health risks’ (Assembly, 2015, p. 17). Goal 4 aims to ‘ensure inclusive and equitable quality education and promote lifelong learning opportunities for all’ (Assembly, 2015, p. 14). Notably, target 4.a stipulates the need to ‘build and upgrade education facilities that are child, disability and gender sensitive and provide safe, non-violent, inclusive and effective learning environments for all’ (Assembly, 2015, p. 17). Target 4.c sets the aim to ‘by 2030,

substantially increase the supply of qualified teachers, including through international cooperation for teacher training in developing countries, especially least developed countries and small island developing States' (Assembly, 2015, p. 17).

These goals on health and education are very much intertwined because of the prominent roles of schools in children and families' lives. When parents have concerns about their children's health, teachers are a key first point-of-contact for information and assistance. Within Britain, 60% of parents whose child had conduct disorder, 47% of parents whose child had emotional disorder, and 70% of parents whose child had hyperkinetic disorder reported that they sought advice or help from a teacher (H. Green, McGinnity, Meltzer, Ford, & Goodman, 2005). These striking percentages highlight the central role that educators play across a range of issues for children's development, including their academic instruction, but also their psychological and behavioural health. As a result, focusing on children's mental health and the role of educators is both globally relevant and necessary in the wider discourse on supporting healthy and educated children so that they can participate fully in society.

This dissertation approaches public mental health from the perspective of intervening through teachers and schools. As such, this chapter reviews the literature around defining social, emotional, and behavioural disorders, teachers' struggles in classroom management, and the additional difficulties in classroom management when including children identified with special educational needs in mainstream classrooms. This background information sets up the two studies which form the dissertation research project: 1) a mixed methods systematic review on effectiveness and experiences of a teacher classroom management programme, and 2) a semi-structured qualitative interview study on exploring the appropriateness and acceptability of implementing a teacher

classroom management programme to support children identified with special educational needs in mainstream classrooms.

Defining and diagnosing social, emotional, and behavioural disorders

Clinical diagnoses of childhood disorders are most frequently decided using the criteria set forth in either the International Classification of Diseases (ICD-10) (World Health Organization, 1992) or the Diagnostic and Statistical Manual of Mental Disorders (DSM-5) (American Psychiatric Association, 2013). Both of these tools set forth a categorical system of classifying disorders. Although there is debate on the philosophical and political merits of using a categorical versus a dimensional classification system, there is a practical need to, at the very least, categorically identify which children need help and what type of help they need (Sonuga-Barke, 1998). It is not for this dissertation to delve into the debate about the relative value in categorical or dimensional classification of childhood disorders. Instead, it is recognised that children's social, emotional, and behavioural health difficulties are generally constructed through categorical classifications to receive clinical support, even if people who interact with the children daily might construct their needs in a more dimensional manner.

This section first outlines the diagnostic criteria used in determining clinical levels of mental and behavioural disorders, since these criteria are a central part of understanding the types of challenges children may experience. Although there is much overlap across the ICD-10 and DSM-5, criteria for a clinical diagnosis diverge on some relevant topics. As such, definitions and diagnostic criteria for relevant mental and behavioural disorders are provided using first the ICD-10 and then DSM-5 manuals, with any discrepancies between the two highlighted. Because this dissertation is interested in school-age children,

the following disorders will be explored in relation to how they manifest in and affect children, rather than adults, when such differences exist.

Conduct disorder. The ICD-10 characterises conduct disorder as ‘a repetitive and persistent pattern of dissocial, aggressive, or defiant conduct’, explaining that such behaviour is ‘more severe than ordinary childish mischief or adolescent rebelliousness’ (World Health Organization, 1992, p. 209). Thus, although the definition implies that children and adolescents demonstrate some mild or isolated acts of aggression or defiance, the presence of these cases alone are not sufficient to diagnose a conduct disorder. Rather, the negative behaviours must be chronic and severe enough to distinguish them from the regular range of behaviours exhibited from peers. Classified in a DSM-5 chapter with disruptive and impulse-control disorders, conduct disorder manifests in behavioural symptoms, such as aggressive or deceitful acts, that ‘violate the rights of others or that violate major societal norms’ (p. 461) and which stem from problems in emotional and behavioural regulation. Oppositional defiant disorder, marked by irritability, vindictiveness, argumentativeness, and defiance, is often a precursor to conduct disorder when present during childhood. Conduct disorder may be diagnosed as early as during the preschool years, although it is often seen initially during middle childhood through middle adolescence (American Psychiatric Association, 2013). It is common for the behaviours associated with conduct disorder to evolve as children mature physically, cognitively, and sexually, and behaviours that ‘emerge first tend to be less serious (e.g., lying, shoplifting), whereas conduct problems that emerge last tend to be more severe (e.g., rape, theft while confronting a victim)’ (American Psychiatric Association, 2013, p. 473).

A number of environmental and genetic factors influence the development of conduct disorder among children. Parental neglect, harsh and inconsistent parenting, physical and sexual abuse, peer rejection, and exposure to violence are all associated with

6

conduct disorder during childhood (American Psychiatric Association, 2013).

Furthermore, having a parent with conduct disorder, alcohol use disorder, bipolar disorder or schizophrenia increases the chances that children will develop conduct disorder (American Psychiatric Association, 2013).

Associations. Conduct disorder often presents as comorbid with attention-deficit/hyperactivity disorder, and the combination of both of these disorders predicts poorer outcomes (American Psychiatric Association, 2013). Children with conduct disorders are often excluded from school (temporarily or permanently), which negatively affects both the amount of teaching time they receive and their likelihood of graduating from secondary school. The DSM-5 notes that particularly ‘in ambiguous situations, aggressive individuals with conduct disorder frequently misperceive the intentions of others as more hostile and threatening than is the case and respond with aggression that they then feel is reasonable and justified’ (p. 472-473). Such hostile attribution bias and any corresponding aggression would not endear children with conduct disorder to others, and peer relationships are likely to suffer.

Hyperkinetic disorder/Attention-deficit hyperactivity disorder. According to the ICD-10, hyperkinetic disorder is evidenced by ‘a combination of overactive, poorly modulated behaviour with marked inattention and lack of persistent task involvement; and pervasiveness over situations and persistence over time of these behavioural characteristics’ (p. 206). Children with diagnosed hyperkinetic disorder display heightened activity levels combined with difficulties controlling their behaviour. These characteristics are not limited to one environment (e.g., school, home, community) and are chronic rather than short-lived. The ICD-10 stresses that, ‘[hyperkinetic] disorders always arise early in development (usually in the first 5 years of life)’ (p. 206). Misbehaviour is not uncommon among children with hyperkinetic disorders, however this is due to

‘unthinking (rather than deliberately defiant) breaches of rules’ (World Health Organization, 1992, p. 206), which is one way of differentiating hyperkinetic disorder from conduct disorder. Thus, hyperkinetic disorders are marked by a general air of impulsivity, deficiency in task persistence, and disproportionately high levels of activity.

Classified as one of the neurodevelopmental disorders in the DSM-5, attention-deficit/hyperactivity disorder (ADHD) is identifiable by ‘a persistent pattern of inattention and/or hyperactivity-impulsivity that interferes with functioning or development’ (American Psychiatric Association, 2013, p. 61). Symptoms must be severe enough to impede children’s abilities to function in daily life or to affect their development, rather than representing levels of attention and activity that would fall within a normal range for children’s developmental level. As such, the DSM-5 notes that, ‘ADHD is most often identified during elementary school years, when inattention becomes more prominent and impairing’ (p. 62). In concurrence with the ICD-10, the DSM-5 also mandates that these symptoms must present in more than one environment (i.e., school and home), while recognizing that manifestations are likely to fluctuate across contexts. Although the DSM-5 concurs with the ICD-10 that ADHD develops during childhood, the DSM-5 requires only that symptoms present prior to age 12, rather than age five. The rationale for this later age marker results from the challenge in retrospectively recalling specific symptomatic details in relation to any earlier age cut-off. This seven-year discrepancy is likely to result in more children meeting diagnostic criteria for ADHD using the DSM-5 than for hyperkinetic disorder using the ICD-10.

Associations. Different poor outcomes are associated with manifestations of the various symptoms of ADHD. The DSM-5 notes that ‘[academic] deficits, school-related problems, and peer neglect tend to be most associated with elevated symptoms of

inattention, whereas peer rejection and, to a lesser extent, accidental injury are more salient with marked symptoms of hyperactivity or impulsivity' (p. 63).

Emotional disorder. Emotional disorders describe feelings of distress, including anxiety and depression. The ICD-10 describes emotional disorders that are diagnosed during childhood as 'exaggerations of normal developmental trends rather than phenomena that are qualitatively abnormal in themselves' (World Health Organization, 1992, p. 210). Thus, emotional disorders as seen in children are typically cases where the emotions are amplified to heightened levels. Examples include separation anxiety disorder, because it is natural for young children to experience some feelings of distress when separated from familiar adults. However, if these feelings are so severe as to negatively impact children's social functioning or if they endure past the early years then children would be considered for diagnosis of an emotional disorder (World Health Organization, 1992).

Prevalence of childhood mental health disorders

The last British Child and Adolescent Mental Health Survey collected data from a nationally representative sample of private households in England, Scotland, and Wales, targeting children between the ages of 5-16-years-old. Results illustrated that nearly 6% (95% CI: 5.1-6.4) of school-age children in these three countries were identified with conduct disorder, nearly 4% (95% CI: 3.2-4.2) with emotional disorder, and 1.5% (95% CI: 1.2-1.8) with hyperkinetic disorder, similar to attention-deficit hyperactivity disorder (H. Green et al., 2005). Beyond Britain, a European review of epidemiological studies reported estimated prevalence rates of 3% (95% CI: 0.0-6.0) for conduct disorder and 5% (95% CI: 1.0-17.8) for hyperkinetic disorder for school-age children between 5-17-years-old (Wittchen et al., 2011).

Similar trends are seen in the United States, where a nationally representative survey of school-age children 8-15-years-old identified over 2% (95% CI: 1.8-2.4) with conduct disorder, almost 9% (95% CI: 7.9-9.3) with attention-deficit hyperactivity disorder, just under 1% (95% CI: 0.5-0.9) with anxiety disorder, and almost 3% (95% CI: 2.1-3.3) with depression (Merikangas et al., 2010). Research from a review of Brazilian studies likewise finds over 4% (95% CI: 2.1-7.9) of children under age 18-years-old identified with either conduct disorder or oppositional defiant disorder (J. Murray, Anselmi, Gallo, Fleitlich-Bilyk, & Bordin, 2013). Therefore, although there are some variations across global context, rates (and confidence intervals) are generally stable and overlapping, indicating that upwards of roughly 10% of children suffer from difficulties across conduct, hyperkinetic, and emotional disorders domains.

However, these prevalence rates are limited to only those children who meet diagnostic criteria for a *disorder*, and they do not account for the potentially many children who have heightened *problems* in any of these domains but are sub-clinical. The Brazilian review established that an additional 20% (95% CI: 15.9-26.9) of children could be classified as having conduct problems (J. Murray et al., 2013), a number which is similarly supported in other countries (Snyder, 2001). In all, then, it is possible that around one-quarter of children experience difficulties with social, emotional, and behavioural health domains. This is not an insignificant problem and as previously highlighted, teachers serve as the most common resource for parents seeking advice about how to support their children who are struggling (H. Green et al., 2005).

Special educational needs and inclusive education

The Salamanca Statement marked the beginning of an international movement towards inclusion of children identified with special educational needs (SEN) in mainstream

schools (UNESCO, 1994). Children might struggle at school because of learning/academic needs, social difficulties, emotional and behavioural difficulties, or frequently combinations of these. In Great Britain, responsibility for their inclusion falls to special educational needs coordinators (SENCOs) and class teachers. Patterns of problematic behaviours, such as noncompliance and aggression, are negatively correlated with children's academic performance (DiLalla, Marcus, & Wright-Phillips, 2004). Similarly, these behaviours are correlated with stressed teacher-pupil relationships and poorer peer interactions (McMahon, Wells, & Kotler, 2006). A comparison of pre-teen children identified with autism spectrum disorders (ASD) to children who showed similar patterns of behaviour at an early age but were not diagnosed with ASD shows that those with the diagnosis had poorer social communication skills and were reported to have more difficulties controlling their tempers (Russell et al., 2012). Longitudinal data following children from school entry at about five-years-old through to age 19 found that those with early identified co-morbid academic and behavioural difficulties were at significantly increased risk of lower reading and mathematics achievement, dropout from secondary school before completion, and use of mental health services (Darney, Reinke, Herman, Stormont, & Ialongo, 2013). Longer-term outcomes include higher likelihoods of being arrested by the police and convicted in court (Collishaw, Maughan, Goodman, & Pickles, 2004). Children identified with SEN, particularly if their needs involve an emotional and behavioural component are a vulnerable group at risk of poor outcomes, whose inclusion is a challenge to mainstream schools.

Defining and diagnosing special educational needs

Traditionally, SEN have been defined from a medical perspective through clinical diagnoses, with a focus on some sort of impairment within the child; however, more

recently and through the inclusion movement the focus is shifting towards how children respond to academic activities and if they are achieving the learning outcomes (Florian, 2008). This section first provides the clinical definitions and then situates SEN within the British context with recent SEN legislation.

Intellectual/learning disability. Historically this condition was termed mental retardation, but such terminology is not widely used anymore. Current practice in British school settings refers to it as learning disability or learning difficulty (not to be confused with specific learning difficulties such as dyslexia, addressed later). Among the academic literature in this field as well as in the DSM-5, the condition is regularly called intellectual disability. The ICD-10 still refers to the condition as mental retardation, but the ICD-11 (which is currently under development and scheduled for publication in 2017) reportedly calls it intellectual developmental disorder. For the purposes of this thesis, the terms intellectual disability and learning difficulty will be used interchangeably to refer to this condition of global deficits in cognitive skills.

Intellectual disability occurs when there is ‘arrested or incomplete development of the mind’ (World Health Organization, 1992, p. 176) and results in decreased intellectual functioning across a number of developmental areas, including motor skills, language, cognitive processing, and social interactions. The ICD-10 clarifies that ‘[adaptive] behaviour is always impaired, but in protected social environments where support is available this impairment may not be at all obvious in [children] with mild [intellectual disability]’ (World Health Organization, 1992, p. 176). Typically, the various levels of intellectual disability are associated with general IQ ranges: mild (50 to 69), moderate (35 to 49), severe (20 to 34), and profound (under 20). The DSM-5 lists essential characteristics of intellectual disability as ‘deficits in general mental abilities (Criterion A)

and impairment in everyday adaptive functioning, in comparison to an individual’s age-,
12

gender-, and socioculturally matched peers (Criterion B)’ (American Psychiatric Association, 2013, p. 37).

Risk/protective factors. Intellectual disabilities typically present in the early developmental stage, with a large variety of aetiologies possible at various stages. Prenatal causes include chromosomal disorders, maternal illness, and drugs or alcohol. Complications during childbirth may also cause intellectual disability. Postnatal aetiologies might include medical problems such as seizure disorders, a traumatic brain injury, ingestion of toxic substances such as lead, or even extreme and sustained social deprivation (American Psychiatric Association, 2013). With a few exceptions among intellectual disorders linked to rare genetic syndromes, intellectual disabilities are not usually progressively degenerative across the lifespan (American Psychiatric Association, 2013).

Associations. Despite the general stability of the symptoms of most intellectual disabilities, this condition is comorbid with numerous other mental health and medical challenges. The ICD-10 states that intellectual disability ‘can occur with or without any other mental or physical disorder...[and] the prevalence of other mental disorders is at least three to four times greater in this population than in the general population’ (World Health Organization, 1992, p. 176). This statistic is similarly reported in the DSM-5, with the most common comorbidities including attention-deficit/hyperactivity disorder, anxiety disorders, autism spectrum disorder, as well as depressive and bipolar disorders. Harmful behaviours among those with intellectual disabilities may be self-directed or aimed at others or property.

Specific speech and language/communication disorders. Specific speech and language disorders can be difficult to diagnose because it is normal for children to ‘vary widely in the age at which they first acquire spoken language and in the pace at which

language skills become firmly established' (World Health Organization, 1992, p. 183). As such, clinicians working with children who demonstrate language difficulties look for: 1) a level of severity in the language delay of at least two standard deviations below the mean, 2) an abnormal pattern (as opposed to patterns that are appropriate for earlier developmental stages), 3) a course that suggests the difficulties maintain across time or deteriorate further, and 4) associated problems such as academic difficulties (World Health Organization, 1992). However, specific speech and language disorders should not be diagnosed when children experience severe deafness, as the speech delays 'are a direct consequence of the hearing impairment' (World Health Organization, 1992, p. 184), or when children's speech deficits are part of a more comprehensive developmental difficulty, such as an intellectual disability.

Specific disorders of scholastic skills and learning disorders. Setting similar criteria as those set forth for diagnosing specific speech and language disorders, the ICD-10 clarifies that specific developmental disorders of scholastic skills are not to be mistaken for low academic attainment resulting from a dearth of learning opportunities. Rather, it is hypothesized that these disorders 'stem from abnormalities in cognitive processing that derive largely from some type of biological dysfunction' (World Health Organization, 1992, p. 188). Unlike the development of speech and language, which generally follows a developmental maturation course, the academic skills that would be found lacking to meet diagnostic criteria for this disorder are not naturally gained through maturation. Since these skills must 'be taught and learned' and given that '[countries] vary widely in the age at which formal schooling is started, in the syllabus followed within schools, and hence in the skills that children are expected to have acquired by different ages' (World Health Organization, 1992, p. 189), it is difficult to operationalize this disorder in such a way as to have validity across countries.

Pervasive developmental disorders. The ICD-10 describes pervasive developmental disorders as ‘characterized by qualitative abnormalities in reciprocal social interactions and in patterns of communication, and by restricted, stereotyped, repetitive repertoire of interests and activities’ (p. 198). Therefore, these disorders are not identified by cognitive deficits but by the distinguishing behaviours displayed, which may vary in the degree to which they manifest across environments but nevertheless must remain ‘a pervasive feature of the individual’s functioning in all situations’ (World Health Organization, 1992, p. 198).

Special educational need (SEN). Current statutory guidance from the UK Department for Education defines the criteria for establishing children with special educational needs and disabilities as well as outlines the special educational provision for children aged 0-25, to which they are legally entitled. According to the 2014 Code of Practice, children with SEN have either:

1. a learning difficulty that results in significantly increased struggles to engage in academic material as compared to most others their age, or
2. a disability that negatively impacts their ability to interact within mainstream schools. (Department for Education & Department of Health, 2015)

Special educational provision for children over age two in school is described as ‘educational or training provision that is additional to or different from that made generally for other children or young people of the same age’ (Department for Education & Department of Health, 2015, p. 5). A disability is legally defined in this context as a physical or mental impairment, which significantly and negatively impacts daily life activities for at least one year (Department for Education & Department of Health, 2015; Legislation.gov.uk, 2010).

Here we see how, for children with clinical levels of any of the above social, emotional, and behavioural disorders and well as other physical impairments, it is the need for additional support to access the mainstream academic curriculum, rather than something inherent within themselves, that leads to a designation of having a special educational need. Thus, it is not enough to demonstrate that children have a clinical disorder; they must also require assistance above and beyond what is regularly provided for general instruction to receive a classification of SEN. Similarly, children without clinical diagnoses of any of the above disorders may receive a classification of SEN if they require additional support outside of the general instruction to access appropriate educational outcomes.

Prevalence of children identified with SEN and their trajectories

Children identified with SEN experience a number of disadvantages affecting all areas of their lives, including academic performance, mental health, social participation, and socio-economic stability. There are many children identified with SEN in Great Britain: data from the Millennium Cohort Study (Parsons & Platt, 2013) indicate that 13% of children aged seven years are identified by either a parent or teacher as having SEN, with an additional 4% receiving a Statement of Needs. Almost half of those children identified with SEN present with at least two different types, including speech, language and communication, learning difficulties, and emotional-behavioural difficulties (Parsons & Platt, 2013).

As previously mentioned, the community-based British Child and Adolescent Mental Health survey identified psychiatric disorders among 10% of school-age children: nearly 6% with a conduct disorder, nearly 4% with an emotional disorder, 1.5% with a hyperkinetic disorder (H. Green et al., 2005). Around 1% of children in this survey were

identified with ASD (H. Green et al., 2005). Notably, 52% of children with conduct disorders, 35% of children with emotional disorders, and 71% of children with hyperkinetic disorders were identified with SEN, which indicates the close relationship between poor mental health and difficulties in coping at school. Children identified with SEN are a heterogeneous group, and research has shown that different disadvantages are associated with the various types of SEN. Particularly, 'children identified with learning, behaviour or speech difficulties experience the most socio-economic disadvantage, in terms of parental education...and income poverty...Children with behaviour difficulties are also more likely to be in a lone-parent household' (Parsons & Platt, 2013, p. 17).

The developmental trajectory of children with SEN differs from their peers who are not identified with SEN. Between the ages of three to seven years, children identified with SEN experience increased peer problems, hyperactivity, and emotional difficulties in contrast to children's typical trajectories, which tend to show decreases in peer problems and hyperactivity over this same period (Fauth, Parsons, & Platt, 2014). While all children tend to experience increases in emotional problems across these early school years, children identified with SEN have the double disadvantage of starting with higher levels of emotional problems at age three and experiencing more rapid escalation of these problems (Fauth et al., 2014). Children identified with SEN have been found to experience a cumulative risk effect, whereby with each additional risk factor to which they are exposed the chance of later developing behavioural difficulties is heightened exponentially (Oldfield, Humphrey, & Hebron, 2015). They are, therefore, a highly vulnerable group and intervention at school may be an important determinant in stemming this poor trajectory.

Teachers' difficulties with classroom management

Despite their central role in the day-to-day lives of school children as well as their position of importance for parents, teachers regularly report feeling underprepared for the reality of managing a classroom, especially classrooms that include children identified with SEN. This is neither a recent phenomenon nor is it limited to newly qualified teachers, although much research has focused on the difficult transition from theory to practice for new teachers. Wickman (1928) surveyed teachers and mental health clinicians, finding that teachers were very concerned about children's externalising behaviours whereas clinicians were more concerned about the internalising issues. Wickman argued that teachers were in need of better education that would focus on, among other things, how to address the emotional disturbances underlying outwardly observable problematic behaviours of children. Nearly six decades later, difficulties with classroom discipline still topped teachers' lists of most serious problems (Veenman, 1984). This concern has persisted into the new millennium, with teachers continuing to report concerns about their inability to manage difficult behaviours in the classroom (Meister & Melnick, 2003). Kagan (1992) concluded that teachers possess preconceptions about the classroom that do not match reality and are unequipped with appropriate strategies for classroom management. As a result, these teachers often rely upon authoritarian strategies as a way of discouraging problematic behaviours (Kagan, 1992).

It was out of this combined need to support the social, emotional, and behavioural development of children in academic environments as well as to properly equip teachers with effective strategies for managing daily classroom activity that research emerged in the area of classroom management. According to Evertson and Weinstein (2013), classroom management is the set of 'actions teachers take to create an environment that supports and facilitates both academic and social-emotional learning' (p. 4). It is a multifaceted

construct that includes both preventative (e.g., routines) and reactive (e.g., consequences) strategies (Korpershoek, Harms, de Boer, van Kuijk, & Doolaard, 2014). The two dominant historical movements in researching classroom management stemmed from behaviourist (e.g., shaping behaviours) and ecological (e.g., analysing the classroom environment) approaches (Brophy, 2013). As research into specific strategies of classroom management grew, programmes were developed that incorporated these components. Korpershoek et al. (2014) identified four primary purposes for classroom management programmes: 1) changing teachers' behaviours, 2) improving the teacher-child relationship, 3) changing children's behaviours, and 4) improving children's social-emotional development. Korpershoek et al. (2014) are clear in asserting that these categories 'are not mutually exclusive' (p. 646). Therefore, classroom management programmes may incorporate elements across these four foci.

Although there are a variety of programmes from which to choose, the focus of this dissertation is the Incredible Years Teacher Classroom Management (IY TCM) programme. Given the international focus on children's health and education (Assembly, 2015) as well as the importance of teachers and schools as a first point-of-contact for children and families (H. Green et al., 2005), determining the overall effectiveness of a programme like IY TCM has implications for how society addresses social, emotional, and behavioural health difficulties among school-age children. Moreover, in understanding the core mechanisms and key contextual features of successful IY TCM implementation, decision-makers are armed with deeper knowledge of how they might expect this programme to benefit both children and educators as well as insights into any adaptations needed to respond to contextual features unique to their area. Considering the commitment to investing in infrastructure to widen access to education at all levels (early childhood through secondary) and the systems in place to train teachers to become

effective educators (Assembly, 2015), implementing a teacher training programme as both a professional development opportunity for educators and a public health intervention for children has the great potential to be an effective use of resources and appropriate avenue by which to reach a large number of children.

Incredible Years Teacher Classroom Management Programme

Developed by Webster-Stratton, the Incredible Years Series is a suite of training programmes aimed at parents, teachers, and children to decrease children's aggression and conduct problems, increase children's prosocial behaviours, and increase parent/teacher competencies (The Incredible Years, 2013). The IY TCM programme was developed to address problematic behaviours via training teachers to use positive and proactive management strategies. Although this programme was originally designed for a clinical group of children with diagnosed oppositional defiant disorder and conduct disorder, is also beneficial when implemented with wider groups of children from community samples (Webster-Stratton, Reid, & Hammond, 2001). It involves six learning components:

1. The importance of teacher attention, encouragement, and praise
2. Motivating children through incentives
3. Preventing problems—The proactive teacher
4. Decreasing inappropriate behaviour
5. Building positive relationships with children
6. Social skills and problem solving training in the classroom (Webster-Stratton, 2001).

Two trained facilitators lead group sessions for small groups of teachers over the course of six, monthly full-day sessions (or 18-20, weekly two-hour sessions). This is a highly manualised programme, and facilitators are required to attend training in the

techniques and to be certified to deliver the course to groups of teachers. The programme incorporates a variety of strategies to introduce, discuss, and practice the strategies. This includes showing video vignettes of teachers in full-size classrooms (28 children) and of special education classrooms with smaller teacher-to-child ratios. Teachers watch the vignettes and use the scenarios presented to spark dialogue and problem-solve. Another important training technique involves the use of role play, where the facilitators and teachers act out scenarios that the teachers might expect to experience in their classroom. These offer teachers the opportunity to practice using learned strategies and become comfortable (or at least more familiar) with what to say and how to act, before they apply the technique in their classroom during real-world situations. In between the monthly sessions, teachers implement the newly learned strategies in their classrooms and reflect on their successes and struggles individually. At the following training session, members discuss their experiences with the new strategies to allow for the group to support each other, learn from one another, and problem solve collaboratively before once again heading back into their classrooms (Webster-Stratton, 2001).

Theory of intervention

IY TCM is based primarily on Patterson's childhood aggression theory, Bandura's social learning theory, and Bowlby's attachment theory. Patterson described a developmental model of antisocial behaviour, with three chronological steps during children's development resulting in the development of antisocial behaviour. Translating Patterson's theory from the home to the school environment, these steps are: 1) ineffective teaching leading to conduct disorders, 2) conduct disorders leading to academic failure and peer rejection, and 3) involvement in deviant peer groups further accelerating the development of antisocial behaviour, becoming chronic (Patterson, DeBaryshe, & Ramsey, 1989;

Patterson, Reid, & Dishion, 1992). Patterson's coercion hypothesis argues that teachers reinforce negative and coercive behaviours in children by providing attention to the negative behaviours, approving the negative behaviour, or (most importantly) allowing the coercive behaviour to lead to avoidance of an aversive behaviour. Thus, when children want to avoid a situation, they use negative behaviours as a form of coercing their teachers to remove the situation, a process that conditions children to realise the benefits of negative behaviours. Patterson warns that over time, the intensity of the coercion process will escalate.

However, it is not just the presence of negative and coercive behaviours but also the lack of positive reinforcement of prosocial skills that risks development of conduct problems and antisocial behaviour (Snyder & Patterson, 1995). By ignoring or otherwise responding inappropriately to prosocial behaviours of children, teachers can condition children's negative behaviours. The IY TCM programme applies the developmental model of antisocial behaviour to the teacher-child relationship. It aims to educate teachers on how to more appropriately respond to children's attempts at coercive behaviours and to their displays of prosocial behaviours in a school context.

Bandura's social learning theory posits that neither internal nor external forces completely account for people's actions. Instead, people are constantly involved in a reciprocal interaction between themselves and their environment (Bandura, 1986; Bandura & McClelland, 1977). Bandura asserts that people develop new behaviour patterns through direct and indirect learning. From this perspective, observation of others is just as important as individual experience. People create their own opinions on what behaviours are successful and acceptable based on a feedback loop of information from their direct and indirect experiences (Bandura & McClelland, 1977). The IY TCM programme utilizes this theory of social learning both in the process by which teachers are educated on

classroom management strategies (e.g., videotape vignettes, role-play) and in the shaping of children's behaviours in the classroom (e.g., experiencing or seeing another child's prosocial behaviour being rewarded).

Finally, IY TCM translates the relationships described in Bowlby's attachment theory to the teacher-child relationship. Bowlby's attachment theory posits that infants' attachments with their mothers are an essential influence on how those infants later interact with others and the ways in which they shape their relationships (Bowlby, 1958, 1960, 1969). Bowlby discusses the bilateral direction of influence in the adult-child relationship, suggesting that children affect adults just as adults affect children. Ainsworth expounded on Bowlby to list different classifications of attachment, particularly secure and insecure (Ainsworth, 1985; Ainsworth, Blehar, Waters, & Wall, 1978). Thus, with IY TCM, the goal is to shape positive teacher-child relationships such that young children in school develop secure attachments to their early childhood and primary school teachers, because these attachments reflect how the children later interact with others in the school setting. Researchers in the field assert that, 'the essence of a successful preschool resides in the quality of the child-teacher relationship and the ability of teachers to provide a positive, consistent, and responsive environment' (Webster-Stratton, Reid, & Hammond, 2001; Bear, Webster-Stratton, Furlong, & Rhee, 2000). Thus, IY TCM combines childhood aggression theory, social learning theory, and attachment theory in applying the core tenets of these models to the school environment and the teacher-child relationship. Ultimately, by intervening at the teacher-level to change teachers' behaviours in classroom management strategies, IY TCM aims to affect children's behaviours and improve classroom relationships.

Mixed methods research

Research has often been described as quantitative (dealing in numbers) or qualitative (dealing in words), suggesting a clear dichotomy between two different and opposing approaches. Reichardt and Cook (1979) called on researchers to reframe the qualitative-quantitative debate in a more positive light. They asserted that the debate was ‘obscuring issues and unnecessarily creating schisms...when it should be building bridges and clarifying genuine disagreements’ (Reichardt & Cook, 1979, p. 9). They argue that while paradigmatic stance is important, it should not be ‘the sole determinant of the choice of methods’ (Reichardt & Cook, 1979, p. 16). Rather, the specific research question for a study should also serve as an important determining factor in selecting appropriate methods (Reichardt & Cook, 1979).

Indeed, the knowledge derived from qualitative research and that from quantitative research can be complementary rather than oppositional (Greene & McClintock, 1985; Reichardt & Cook, 1979). Greene and McClintock (1985) explain this as ‘one set of results enriching, expanding upon, clarifying, or illustrating the other’ (p. 524), the purpose of which is between-method triangulation.

Rossmann and Wilson (1985) make the case for moving beyond this situationalist stance of viewing the two approaches as complementary yet distinctly separate in favour of a pragmatist stance, which allows for within-study integration of both qualitative and quantitative methods. They argue that both qualitative and quantitative methods may be used to corroborate, elaborate, and initiate the other. By this, they explain that collaboration serves to ‘confirm or disconfirm established results’ (p. 634), ‘elaboration provides richness and detail’ (p. 636), and initiation ‘suggests areas for further exploration, or recasts the entire research question’ (p. 637). Their conclusion is that ‘numbers and words can be used together in a variety of ways to produce richer and more insightful

analyses of complex phenomena than can be achieved by either one alone' (Rossman & Wilson, 1985, p. 641).

Indeed, Teddlie and Sammons assert that applying mixed methods research to the field of educational effectiveness is valuable for many reasons, including the integration of both testing theory and discovering new knowledge as well as in finding both convergent and divergent results, leading 'to a re-examination of the conceptual frameworks and the assumptions underlying each of the two components' (Teddlie & Sammons, 2010, p. 125). This dissertation seeks to do just this – to re-examine the theoretical understanding of how the IY TCM programme operates and is experienced by people and to illuminate the many ways by which it can be judged effective (or not). Due to the additional insights possible when approaching research from a pragmatic paradigm and in applying mixed methods, it is contended that such studies might be more valuable to both policy-makers and practitioners (Sammons, 2010). Indeed, an aim of this dissertation is for not only practitioners and policy-makers to find the insights and conclusions of this study valuable, but also for researchers to see the importance of conducting mixed methods systematic reviews and to set future research agendas that prioritise both quantitative and qualitative evidence in research syntheses.

Study One

The first study in this dissertation is a mixed methods systematic review of the IY TCM programme. This review utilises multilevel meta-analysis, qualitative meta-synthesis, and cross-synthesis to draw conclusions and make recommendations using all of the available quantitative and qualitative evidence.

Evidence-based education. Approaching research from a mixed methods perspective is based not only in pragmatism but also in Sackett's definition of evidence-

based medicine, which has been since applied to the social sciences, including public health and education intervention research. Sackett describes evidence-based medicine as:

'[...] the conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual patients. [...] Good doctors use both individual clinical expertise and the best available external evidence, and neither alone is enough. Without clinical expertise, practice risks becoming tyrannised by evidence, for even excellent external evidence may be inapplicable to or inappropriate for an individual patient. Without current best evidence, practice risks becoming rapidly out of date, to the detriment of patients.' (Sackett, Rosenberg, Gray, Haynes, & Richardson, 1996, pp. 71-72)

Drawing a direct parallel from medicine to education, teachers are the highly trained and skilled professionals of interest, rather than doctors. Thus, the call becomes one for teachers to practice evidence-based education by incorporating both academic research and experiential data to implement appropriate and effective teaching strategies in the classroom. Following from this, any review of the literature should consider evidence from experimental studies designed to establish causal relationships (such as randomised controlled trials) and from qualitative studies designed to understand people's experiences of and views about the programme. Doing so serves to more comprehensively summarise and interpret the relevant data on any given topic – such as is done in this dissertation on the IY TCM programme – in accordance with an evidence-based approach that values both research evidence and practitioner expertise. The findings are in turn of more help to decision-makers, who will be able to take into account how the intervention might be received and how it will perform in their context, knowing how it was received and the results it produced in different types of other contexts.

Aims and research design. To this end, Study One aims to illuminate the effectiveness and experiences of the IY TCM programme. To accomplish this aim, this systematic review answers three overarching questions:

1. What are the effects of the Incredible Years Teacher Classroom Management programme on teacher and child outcomes?
2. How do people experience/perceive the programme and its effects?
3. When considered together, what conclusions can we draw based on the effectiveness and experiential data?

There are many different approaches to conducting mixed methods studies, into which this discussion will not deviate except to direct interested readers to Teddlie and Tashakkori (2009) and Creswell and Plano Clark (2011). This study is conducted according to a mixed methods convergent parallel design (see Figure 1.1), which involves a phase in which qualitative and quantitative data are concurrently yet independently collected and analysed, followed by a phase which sees these two strands joined together and the data merged for subsequent interpretation and discussion (Creswell & Plano Clark, 2011; Teddlie & Tashakkori, 2009). Based in pragmatism, the convergent parallel design places the mixed methods nature of the research question at the centre, casting aside any controversies arising from paradigm differences between constructivism and positivism in favour of applying a practical approach to research (Sammons, 2010). This design is most appropriate to answer the research question at hand because it allows for the synthesis of corresponding data across qualitative and quantitative results so as to more fully illuminate the phenomenon of interest, valuing each strand equally without placing one above the other.

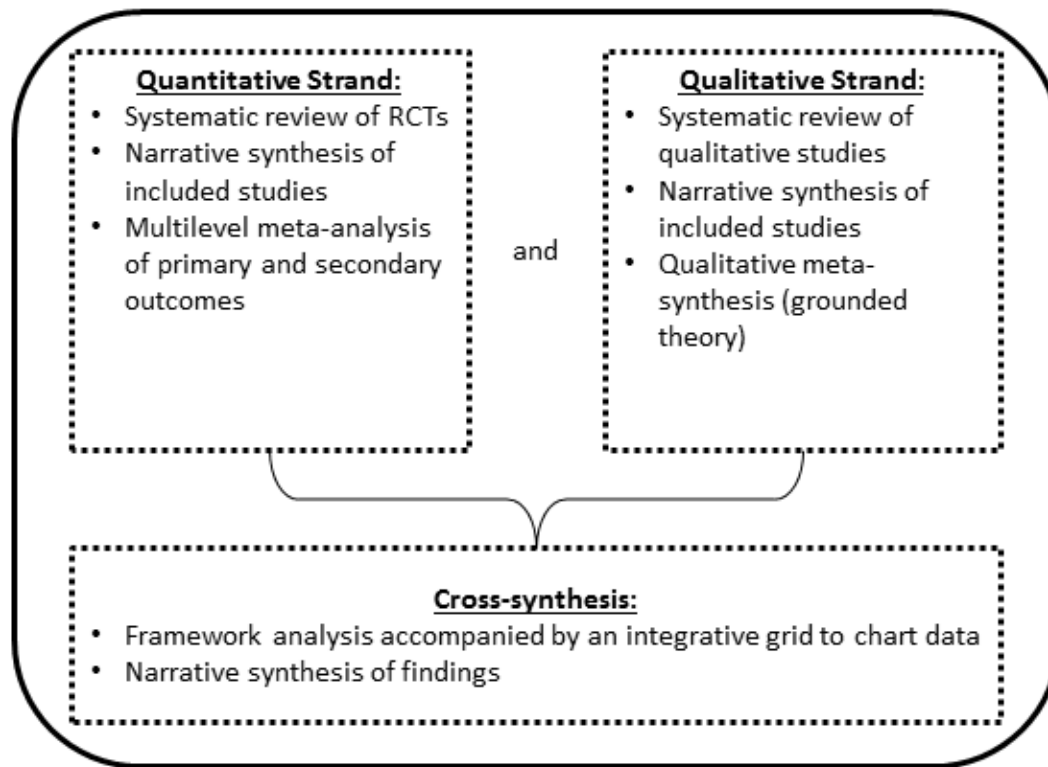


Figure 1.1 Illustration of the convergent parallel mixed methods design for this study

Novel contribution. This review is distinct from two broader reviews conducted within the past five years, both of which warrant attention. In the first review, researchers examined universal teacher classroom management programmes on child disruptive and aggressive behaviours across primary and secondary school (Oliver, Reschly, & Wehby, 2011). Although this review was focused on school, it did include studies that contained additional intervention components, such as parenting. For their meta-analysis, Oliver and colleagues addressed the issue of studies reporting multiple effect sizes for child disruptive or aggressive behaviour by calculating the study average and entering this average into the meta-analysis. Furthermore, due to differences between studies reporting data at individual child level and clustered classroom level, this review chose to transform child level data to the classroom level. They concluded that the classroom-level effect size was

between 0.71 (CI = 0.46-0.96) and 0.80 (CI = 0.51-1.09), corresponding to estimated child-level effect sizes of .22 and .18, respectively (Oliver et al., 2011).

The second recent review studied the effect of training teachers in classroom management strategies on child social, emotional, and behavioural outcomes in primary school (Whear et al., 2012). This review excluded studies that evaluated additional components other than teacher training; however, due to the heterogeneity of included studies, no meta-analysis was conducted in this review. The authors concluded that while no programme had a plethora of high-quality evidence supporting its effectiveness, two programmes stood out as having relatively more evidence to support their implementation: the Good Behaviour Game and the IY TCM programme (Whear et al., 2012).

Although both of these reviews contribute valuable information for understanding the broader area of teacher training in classroom behaviour management on child outcomes, methodological advances and the publication of new empirical studies provide two strong reasons for the need of this current systematic review. Methodologically, this review advances the field by applying multilevel modelling to the meta-analysis, in contrast to the traditional meta-analysis techniques. Whereas the Oliver et al. (2011) review averaged effect sizes within studies before meta-analysing, this review incorporates all of the originally reported relevant effect sizes in the multilevel meta-analysis, therefore retaining as much information as possible and producing a more sophisticated model. Furthermore, due to the heterogeneity between classroom management programmes and study populations, this review specifically targets research evidence available only on IY TCM, one of the two programmes identified by Whear et al. (2012) as having a comparatively more sizable evidence base with which to work. Moreover, by expanding this review's remit beyond that of either of these two previous reviews to include not only trial evidence but also qualitative evidence, this review aims to provide a more

comprehensive understanding of the effectiveness of IY TCM and how it is received by users to inform key decision-makers at various levels of practice and policy.

To this end, this mixed methods systematic review can be divided into three phases: two independent and yet concurrently conducted strands for the quantitative and qualitative evidence and one final cross-synthesis to connect these two strands. The methods for these three phases are detailed in Chapter 2.

Study Two

In their recent systematic review of multiple classroom management programmes, Whear and colleagues concluded that the IY TCM programme was one of only two that had been studied more than once, and of the two it had the more robust evidence-base (Whear et al., 2012). Based on childhood aggression theory (Patterson et al., 1989; Patterson et al., 1992), social learning theory (Bandura, 1986; Bandura & McClelland, 1977), and attachment theory (Bowlby, 1958, 1960, 1969), the IY TCM programme aims to improve teachers' use of positive and proactive management strategies to benefit children's social, emotional, and behavioural development (Webster-Stratton, 2001). Although preliminary studies have been encouraging, there are very few randomised controlled trials that have examined the effects of implementing the IY TCM programme on teachers and their pupils (Baker-Henningham, Scott, Jones, & Walker, 2012; Hutchings, Martin-Forbes, Daley, & Williams, 2013; McGilloway et al., 2010). Qualitative evidence from teachers who have implemented IY TCM strategies suggests that the implementation of these positive classroom behaviour management strategies is more difficult with children identified with SEN in mainstream classrooms (McGilloway et al., 2010).

The mixed methods systematic review reported in Study One of this dissertation concluded that the patterns of effect sizes estimated through multilevel meta-analysis of

randomised controlled trials provide evidence that IY TCM benefits both teacher and child outcomes across multiple countries. It also identified a number of underdeveloped areas surrounding IY TCM through the qualitative meta-synthesis, including potential subgroups of children who are more complex and who might not respond to this intervention. The cross-synthesis highlighted the fact that none of the IY TCM trials has examined possible moderator effects among the subgroup of children identified with SEN. It is also important to consider that SEN provision often differs from regular education provision (e.g., allocation of funding and other resources, training and certification of staff). Therefore, the questions arise of whether the strategies are acceptable to educators working with this subgroup, appropriate for supporting children identified with SEN who also demonstrate problematic behaviours, and whether the intervention is likely to provide additional value above and beyond SEN provision, whatever that might entail within the local context.

Supporting Teachers and childRen in School (STARS) Study. Study Two occurred under an umbrella project entitled: Supporting Teachers And childRen in School (STARS). The STARS Study is a large cluster randomised controlled trial testing the effectiveness of IY TCM in 80 primary schools across the South West Peninsula of England as a public mental health intervention (Ford et al., 2012). However, classroom teachers do not operate in isolation; they are a part of the larger school environment. In Britain, Head Teachers have a duty for ensuring academic provision to the whole school, and school special educational needs coordinators (SENCOs) have a duty to oversee SEN provision for all children identified as having additional needs. SENCOs' responsibilities extend to the provision of training opportunities and guidance to class teachers on the ways to best support their pupils' development (Department for Education & Department of Health, 2015). Because of the intertwined nature of educators' responsibilities and

relationships, the process evaluation that parallels the STARS Study includes focus groups and interviews with the teachers on the IY TCM course, Head Teachers, and SENCOs, among others (Ford et al., 2012; Hansford et al., 2015).

The methods and results reported in subsequent chapters address interviews conducted with school SENCOs on the existing services available to schools when working with children identified with SEN as well as various behaviour management strategies used by educators when supporting children with SEN who also have behavioural needs. These strategies are then compared with IY TCM content to examine the acceptability and appropriateness of IY TCM for educators working with children identified with SEN. Given their leadership role in schools to oversee services to children with identified SEN, the SENCOs are well-placed to provide insight on both their own classroom management experiences as well as on behaviour management strategies encouraged school-wide for other teachers to use when supporting this subgroup of children whose needs are particularly complex.

Aims and research questions. To this end, Study Two follows directly from the results of Study One and aims to lay the preliminary groundwork for determining whether the subgroup of children identified with SEN might be an appropriate target for IY TCM intervention. This study is exploratory in nature and seeks to fill the gap in understanding ‘for whom’ the IY TCM programme is (or is not) appropriate, which could in turn lead to future studies testing the effectiveness in a more targeted manner. To accomplish these goals, this qualitative study answers two research questions:

1. What classroom behaviour management strategies do SENCOs find to be successful when working with children identified with SEN who also have behavioural difficulties?

2. Are the SENCo-generated strategies compatible with the IY TCM content such that the course could be considered acceptable and appropriate for educators who support this subgroup of children?

Novel contribution. This qualitative study contributes new insights to the library of literature on the IY TCM programme through interviewing a relevant group of educators whose perspectives have not been previously sought. The views of educators working with children identified with SEN in mainstream schools are important because, as evidenced in Study One, qualitative research on this programme suggests that SEN (and child complexity more broadly) could be a moderator for treatment effect. In conducting exploratory qualitative research with research questions that are grounded in findings from previous studies in this field, this study fills a gap in the evidence and sets the stage for future studies to test the IY TCM programme in new ways.

STUDY ONE

Mixed methods systematic review of a teacher classroom management programme:
effectiveness and stakeholders' experiences

Chapter 2 : Methods

Introduction

This chapter describes the methods and analysis strategy for all three phases of this systematic review. First, it details the systematic review process for setting the inclusion parameters and search strategy to identify all relevant quantitative, qualitative, and mixed methods studies of the Incredible Years Teacher Classroom Management (IY TCM) programme. It then outlines the data extraction and analysis plan for the quantitative and qualitative strands. Where appropriate, information pertinent to both the quantitative and qualitative strands is presented together, as indicated by subheadings (e.g., see section on search methods). In some cases, information for each of the strands is presented under different subheadings due to the variation in approaches necessary (e.g., see sections on inclusion criteria and analysis plans). Finally, this chapter outlines the methods for cross-synthesising the findings from each of these two strands.

Research questions

A protocol for this review (see Appendix A) was submitted through the University of York's Centre for Reviews and Dissemination PROSPERO database (Nye, Melendez-Torres, & Gardner, 2015).

This review examined the effectiveness of IY TCM and explored people's experiences with the programme. To this end, it answered three intertwined research questions (see Figure 2.1). The quantitative strand addressed the question: What are the effects of the IY TCM programme on teachers' management strategies and children's behaviour, compared to treatment-as-usual, at post-intervention and any measured follow-up time points? The qualitative strand addressed the question: What are the experiences and views of key stakeholders (e.g., teachers, school leaders, parents, children) of the IY

TCM programme (e.g., content, processes, effects)? The cross-synthesis addressed the question: When considered in combination, what can the effectiveness and stakeholders' experience data tell us?

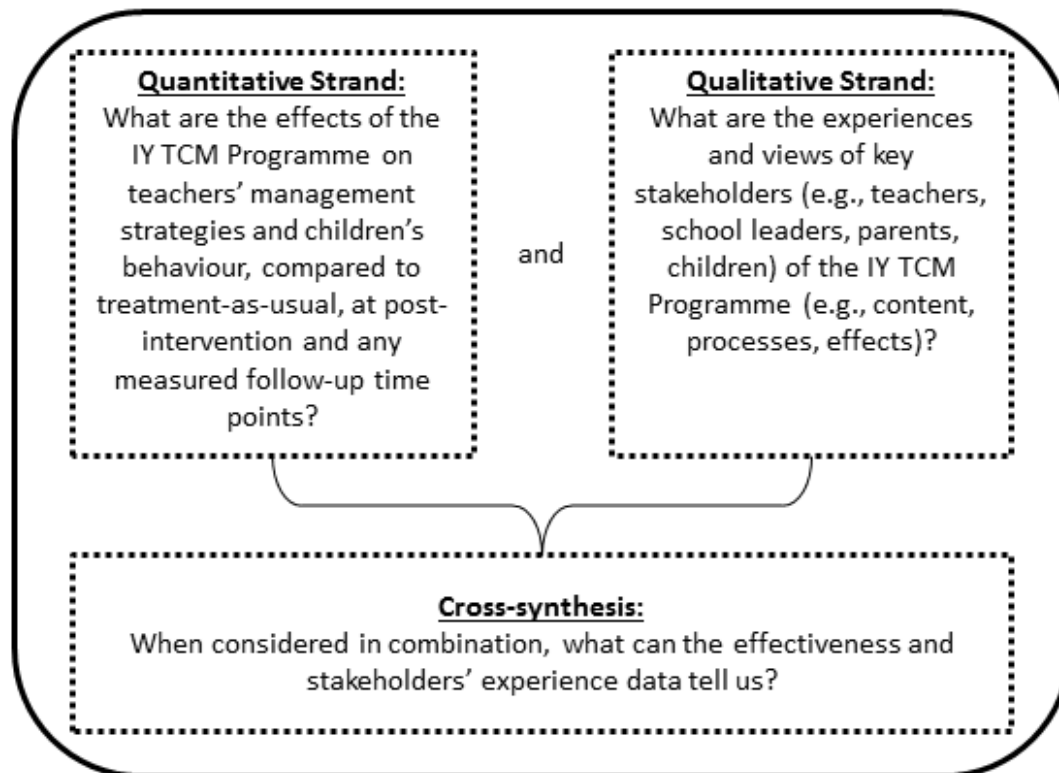


Figure 2.1 Research questions for this mixed methods systematic review

Criteria for Considering Studies for this Review – Quantitative Strand

Types of studies. Studies included in the quantitative strand of this review must have met the following criteria:

1. Population: early childhood and primary school teachers of children ages three- to eight-years-old
2. Intervention: IY TCM
3. Comparison: treatment-as-usual or waitlist control
4. Design: (cluster) randomised controlled trials

Studies that did not meet the above criteria (e.g., delivered IY TCM in conjunction with additional treatment, no control group) were excluded from this review.

Only randomised controlled trials (RCTs) were included in the quantitative strand of this review, because the focus of the quantitative strand was to evaluate the effectiveness of IY TCM through the design with the most rigour for exploring causal links. RCTs randomly assign participants to treatment and control groups, minimising potential selection biases by making the different groups equivalent on both known and unknown variables that might influence intervention outcomes (Fraser, Richman, Galinsky, & Day, 2009). Non-randomised controlled trials might be able to establish baseline similarities on known variables but are unable to do so with unknown variables, which allows for the possibility of systematic distribution of characteristics across groups. In this situation, it is not possible to conclusively determine whether the unknown variables caused the outcomes rather than the intervention treatment. Studies without any control group (e.g., pre-/post-test) are also unable to distinguish whether the outcomes occurred because of the intervention or for some other reason. Most notably, the risk of maturation effects is particularly relevant in studies of child development. For these reasons, included studies for the quantitative strand must have been RCTs, and studies not meeting this criterion were excluded.

It was expected that included studies would be cluster RCTs, which are trials where randomisation occurred at the teacher/classroom or school levels rather than at the child-level. This is because IY TCM is delivered at the teacher level. As with many school-based interventions, there is a natural clustering effect, whereby the organisation of the educational system creates a hierarchy of levels within which interventions are delivered (Goldstein, 1987). Children (level 1) are clustered in classrooms with their teachers (level 2), which are in turn clustered in schools (level 3), which continue to exist within higher-

level units such as local authorities and so forth. Of course, there may be even more levels in between the ones listed here as examples, particularly in large schools that might have multiple classrooms clustered within year groups or multiple teachers per classroom.

When interventions are randomised at the teacher/classroom level, data collected on child outcomes must be adjusted for clustering because it cannot be assumed that the children within a classroom are statistically independent from each other. This process is discussed in more detail in the unit of analysis section.

Types of participants. The participants in the quantitative studies included in this review were early childhood or primary school teachers and children ages three- to eight-years-old. These parameters match the population for which IY TCM was designed.

Types of interventions. The quantitative strand aimed to evaluate IY TCM in isolation from other IY programmes and/or other school-based programmes. The purpose of isolating IY TCM is to determine which outcomes are attributable to the IY TCM programme alone and not in part or in full due to other programmes. Furthermore, the literature on adaptations of interventions warns against both intentional and inadvertent adaptations if not grounded in research. This caution arises because given how little is known about the core components of interventions, it is possible that adapted versions of interventions omit some of the core components through which the programme is theorised to work (Elliott & Mihalic, 2004; Mihalic, 2002). However, cultural adaptations based in research can serve to increase acceptability of the intervention (Lau, 2006) and can support effectiveness among new populations (Griner & Smith, 2006). Cultural adaptations are defined as ‘[...] the systematic modification of an evidenced-based treatment (EBT) or intervention protocol to consider language, culture, and context in such a way that it is compatible with the client’s cultural patterns, meanings, and values’ (Bernal, Jiménez-Chafey, & Domenech Rodríguez, 2009, p. 362). Therefore, studies that evaluated

linguistically or culturally adapted versions of IY TCM were included in so far as they based any such adaptations in research and only adapted the form of the intervention to make it accessible and relevant to the local environment.

Treatment-as-usual is likely to differ across studies and could include professional development training (e.g., In Service Education and Training) or other programmes implemented on a large scale (i.e., local- or national-level initiatives). In order to meet inclusion criteria for this review, any training or programmes given to the control group must have been considered the norm for that area, rather than implemented for the purpose of the research trial. Wait-list controls were also acceptable.

Types of outcome measures.

Primary outcomes. Studies in the quantitative strand reported outcomes measured by independent observers (e.g., members of the research team), teachers (e.g., self-reports or reports of child behaviours), parents (e.g., reports of child behaviours), or other sources (e.g., school records). Including results across these different types of measures allows for balancing the individual weaknesses inherent in each type alone (e.g., teacher-reports are not blind to intervention/control status; observational data is a snapshot and might not be representative of day-to-day behaviours). The primary quantitative outcomes for this review were based on the stated outcomes for the IY TCM programme:

1. Teacher classroom management strategies
2. Child conduct problems (e.g., aggressive or noncompliant behaviours)

In the course of conducting this review and extracting the quantitative data, it became clear that some of the measures of child ‘conduct problems’ included items that aligned better with child hyperactivity (e.g., ‘has difficulty staying seated’) or with child peer problems (e.g., ‘acts bossy with other students’). To ensure the measures included in the meta-analysis of child conduct problems were appropriately grouped, these measures of conduct

problems were excluded from the child conduct problems analysis. Instead, a *post hoc* child behaviour difficulties analysis was added to include not only all measures of conduct problems but also measures of hyperactivity and peer problems, which were reported separately from conduct problems on many measures. In this way, the *a priori* aims of this review were maintained while relevant outcomes were included in a transparently *post hoc* analysis aligning with the way many of the primary studies collected and reported overall child behaviour difficulties. This review included data from immediate trial outcomes (post-intervention) and any follow-up outcomes as available.

Secondary outcomes. In the quantitative strand of this review, secondary outcomes also followed the stated intent of the IY TCM programme and included:

1. Teacher collaboration with parents
2. Child prosocial behaviours
3. Child academic readiness (e.g., reading skills)

Criteria for Considering Studies for this Review – Qualitative Strand

Types of studies. Studies included in the qualitative strand of this review reported qualitative (i.e., non-numerical) data related to the IY TCM programme. More specifically:

1. Data should have come from qualitative studies or from mixed methods studies reporting qualitative outcomes;
2. Studies should have clearly stated that the participants' responses pertained to the IY TCM programme or were part of a larger study on this specific programme—studies did not need to be connected with a trial of IY TCM to be included, but they must have clearly demonstrated direct connection to the IY TCM programme;

3. Those studies that combined the IY TCM programme with any other additional programme (i.e., not considered ‘treatment-as-usual’ for target population) were included so long as data specific to the IY TCM programme were reported separately. (Had data not been isolated in reports, separate analyses would have addressed qualitative data from studies that combined the IY TCM programme with additional components.)

Studies were excluded if data reported only on non-Incredible Years classroom management or teacher training programmes or if the qualitative data reported on either the IY Parent Training or Child Training programmes implemented without the IY TCM programme.

Types of participants. Participants in the qualitative studies were ‘key stakeholders’, and could have included those directly involved with the IY TCM programme (e.g., class teachers, group facilitators) or those more indirectly involved (e.g., head teachers, teaching assistants, parents, children).

Types of findings. The qualitative strand in this review explored social processes and sought to understand the interactions of key stakeholders with the content and delivery of the IY TCM programme, as well as their subsequent views of the strategies and implementation of the techniques in their practice. Qualitative findings might have been based on data collected during or after IY TCM implementation. Any data collected before implementation was only included if it directly addressed the IY TCM programme (as opposed to classroom management in general or other topics), because this review was interested in stakeholders’ experiences and views on IY TCM specifically.

Search Methods for Identification of Studies – Quantitative and Qualitative Strands

One systematic search was conducted for studies to include in both the qualitative and quantitative strands. To identify relevant studies, this review searched electronic databases, websites related to the Incredible Years Series, and reference lists of included studies (see Figure 2.2). The search also extended to studies recommended by researchers in the field, who received emails asking if they could identify any additional studies that were currently on-going, in press/not yet published, unpublished, or otherwise unlisted in electronic databases or online. All researchers who published papers regarding the IY TCM programme or related topics were included in these emails. No studies were identified solely from communicating with researchers in the field. All researcher-identified studies were already identified from database and website searches. Reference lists of included studies were searched for any relevant studies but no additional studies were found solely through this method, either.

Electronic searches. To increase specificity without compromising the sensitivity of any electronic database searches, the search terms focused on identifying IY TCM, which is a branded programme that operates according to a manual. Initial search terms were first piloted in 2013 for a master's thesis (Nye, 2013) and the search terms were pared down to the most parsimonious set without compromising sensitivity (see Figure 2.3).

Before administering IY TCM, facilitators must attend training for certification (The Incredible Years, 2013). According to the Incredible Years website, research on the effectiveness of any of the IY Series may only be undertaken if group facilitators have completed IY certification. This information, combined with the strategy of searching 'all fields' (rather than just key words, titles, and/or abstracts) in electronic databases, suggests

that limiting the search terms to variations of the intervention name did not compromise the integrity of the search output.

Databases:	
1.	Applied Social Sciences Index and Abstracts (ASSIA)
2.	British Education Index
3.	Cochrane Central Register of Controlled Trials (CENTRAL)
4.	CINAHL
5.	EMBASE
6.	ERIC
7.	Linguistics and Language Behaviour Abstracts
8.	MEDLINE
9.	PAIS International
10.	ProQuest Dissertations and Theses
11.	Social Services Abstracts
12.	Sociological Abstracts
13.	PsycINFO
Websites:	
1.	Campbell Collaboration Library http://www.campbellcollaboration.org/lib/
2.	Cochrane Collaboration Library http://www.cochranelibrary.com/cochrane-database-of-systematic-reviews/
3.	Incredible Years Series Library http://incredibleyears.com/research-library/
4.	Centre for Evidence-Based Early Intervention https://www.bangor.ac.uk/psychology/cebei/publications.php.en

Figure 2.2 Electronic searches

1	"incredible years".af.
2	"program*".af.
3	"teacher classroom management".af.
4	"teacher training".af.
5	2 or 3 or 4
6	1 and 5

Figure 2.3 Search strategy

Data Collection – Quantitative and Qualitative Strands

Selection of studies. Studies identified from electronic searches and personal contacts were combined, and duplicates removed. Two independent reviewers screened the titles and abstracts for potential relevance to this project. Potentially relevant reports were retrieved in full-text and assessed for eligibility using predetermined criteria:

- Quantitative criteria:
 - Study design – RCT?
 - Population – early and primary school teachers and children ages three- to eight-years old?
 - Intervention – IY TCM?
 - Comparison – treatment-as-usual or waitlist control?
- Qualitative criteria:
 - Study design – qualitative (e.g., interviews, focus groups)?
 - Population – direct or indirect stakeholders?
 - Time – during or after training and implementation of IY TCM?

In cases of uncertainty over the status of any studies, additional information was sought from the study authors, and the reviewers reached consensus before proceeding.

Data Extraction and Analysis – Quantitative Strand

Extraction. Quantitative data were extracted into Microsoft Excel 2010 using a uniform system and independently double-checked by a second reviewer. In cases where data were unclear or unreported, attempts were made to contact the study authors for clarification. Data extraction included information on the study number, all reported outcomes (e.g., conduct problems, prosocial behaviour), the mode of data collection (e.g., teacher, observer, parent), measure used (e.g., Strengths and Difficulties Questionnaire, Teacher-Pupil Observation Tool), raw scores for both intervention and control groups (N, mean, SD), the intraclass correlation coefficient (ICC), and whether the item needed to be reverse coded (e.g., compliance, peer problems). Where the ICC was reported for the whole sample (e.g., all observed children), but not for other reported subgroups (e.g., high- or low-risk children), it was assumed that the ICC for these subgroups was the same as for the whole sample in that study. Where the ICC was not stated in a study, the mean ICC on that specific outcome was calculated from the other included studies and used.

Risk of bias. All studies included in the quantitative strand were assessed for risk of bias. In situations where multiple reports (e.g., both a published journal article and PhD dissertation) were available for an individual study, information available across all reports was considered when making decisions pertaining to risk of bias.

Quantitative studies were assessed using the Cochrane ‘Risk of Bias’ tool. In situations such as lack of participant blinding (not possible with this intervention) or insufficient details regarding any of the different domains, a standard rating of unclear risk was assigned, following the directions outlined in the Cochrane Handbook (Higgins & Green, 2008). Any studies found to be at high-risk for bias would have been separated out in sensitivity analyses to determine any undue effect on overall outcomes.

Measures of treatment effect. The quantitative strand of this review examined child and teacher outcomes as intended from the IY TCM programme objectives and logic model. Studies were not excluded based on reported or unreported outcomes or on the method by which outcomes were measured. Thus, studies that examined additional or different teacher and child outcomes were included, at minimum, in the narrative synthesis of this quantitative strand. In such cases, data not related to this review’s primary and secondary outcomes were omitted from any multilevel meta-analysis.

Unit of analysis issues. Operating from the logical assumption that the hierarchical structure within schools precludes independence at any given level (i.e., children in one classroom are likely to be more similar to each other than to children in another classroom etc.; see Figure 2.4), outcome data evaluating the effectiveness of any intervention implemented in the school setting must account for its nested nature (Bateman & Jones, 2003; Goldstein, 1987; Goldstein, Yang, Omar, Turner, & Thompson, 2000). In school-based studies, it is also common to collect data from the same child across multiple time points, in which case the data at each stage is nested hierarchically and non-independently within each child.

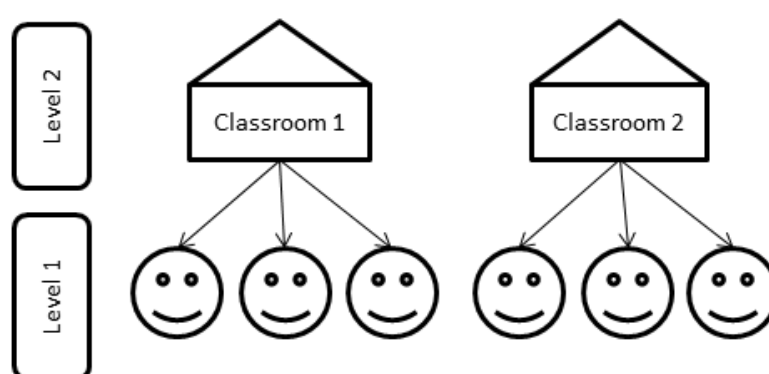


Figure 2.4 Multilevel analysis in primary studies

In empirical studies, multilevel models allow analysts to account for the inherently hierarchical and plausibly non-independent nature of the data. Multilevel models accomplish this task by modelling residual variance of outcomes at a lower level in relation to both the effect of higher levels on the lower level outcomes as well as to any unexplained error (Bateman & Jones, 2003). This is important because large intraclass correlations subsequently result in over-rejection of the null-hypothesis for under-estimating standard errors (Bateman & Jones, 2003; Goldstein, 1987; Goldstein et al., 2000). It was expected that studies included in the quantitative strand of this review would apply multilevel analyses or otherwise account for the clustering within their data.

To account for clustering effects with continuous data, researchers must adjust the sample size to narrow the standard error, which serves to protect against Type I Error (when significant effects are found erroneously). The equation used to adjust the sample size is:

$$n_a = \frac{n_r}{de}$$

Where n_a is the adjusted sample size, n_r is the raw sample size, and de is the design effect, calculated by:

$$de = 1 + (M - 1)ICC$$

Where M is the average cluster size, and ICC is the intraclass correlation coefficient.

Statistical analysis. In the context of a systematic review and meta-analysis, which aggregates quantitative findings across primary studies, it is not uncommon for individual studies to report data from multiple measures for an outcome (e.g., teacher-reported and parent-reported child behaviour). In this way effect sizes are clustered within studies, much like children are clustered within classrooms. The reason for applying

multilevel modelling techniques to the meta-analysis is *not* because of the clustered design utilised in the individual empirical studies. Rather, it is because of the clustered nature of the effect sizes within each study in this review (see Figure 2.5). The issue of data dependency in multilevel meta-analysis is one of dependent effect sizes, which must be accounted for in any analysis to avoid overestimating effect sizes through underestimating standard errors (Van den Noortgate, López-López, Marín-Martínez, & Sánchez-Meca, 2013; Van Den Noortgate & Onghena, 2003). Applying multilevel analysis techniques to a meta-analysis allows researchers to approximate the overall effect size (grand mean) on an outcome using all of the relevant data available. In traditional meta-analyses, researchers might average across effect sizes within a study to create one effect size per study, as was done in the Oliver et al. (2011) study, or they might split the analyses by specific outcome (e.g., mode of data collection), as was done in the Nye (2013) study. However, while both of these approaches to meta-analysis are common, they introduce limitations. Averaging the effect sizes within a study does not appropriately account for how they might be correlated due to their nested nature or take into account variance and error at the different levels. Furthermore, the splitting approach (selecting only one effect size per study for each outcome) introduces researcher bias due to judgement decisions as to which effect size might be best to select (Polanin, 2013). Multilevel meta-analysis avoids these limitations, producing a more sophisticated estimate of the overall effect of an intervention on the target outcome.

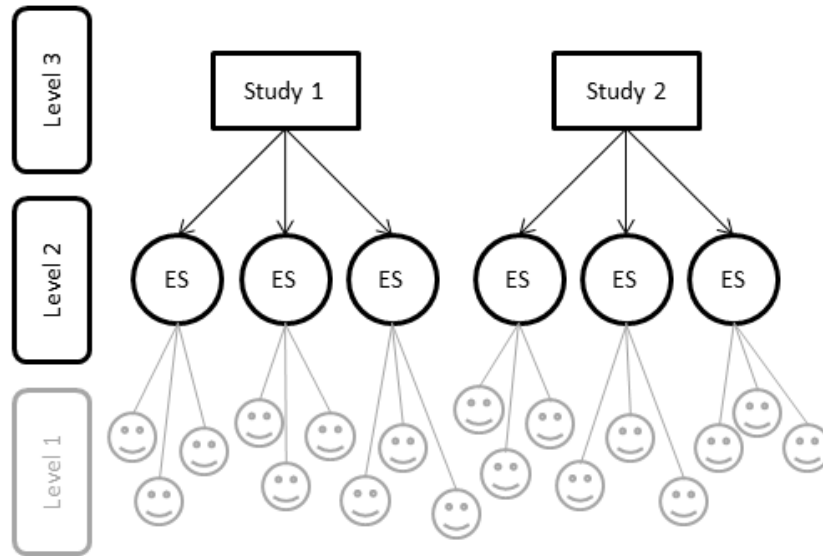


Figure 2.5 Multilevel meta-analysis with first level implied

In this multilevel meta-analysis, the first level (level 1) is implied, because it consists of the individuals whose data is included in the calculations for effect sizes. The second level (level 2) consists of the effect sizes for each outcome of interest, which provide a within-study level intervention-control comparison. The third level (level 3) represents each included study, which provides a between-study comparison. Since each study (level 3) reported many effect sizes (level 2) measuring the same constructs (e.g., child compliance, noncompliance, conduct problems), using a multilevel meta-analysis instead of a traditional meta-analysis provides the best summary estimate (grand mean) of the effectiveness of the intervention compared to control. The equation for a multilevel random-effects model is:

$$T_{ij} = \beta_{00} + \mu_{(2)i} + \mu_{(3)} + e_{ij}$$

Where T_{ij} is the observed effect size, β_{00} is the grand mean effect size, $\mu_{(2)i}$ is the variance across effect sizes within study i (level 2), $\mu_{(3)}$ is the variance across all included

studies (level 3), and e_{ij} is the fixed-effect variance for the j th effect size (level 2) within the i th study.

Using the raw data reported in each included study, adjustments were first made according to the above ICC equations for the clustering of participants within studies (i.e., children within a classroom) before calculating the Hedges' g effect sizes (level 2) and running these in the multilevel meta-analysis to calculate the Hedges' g grand mean (β_{00}). Hedges' g was calculated instead of Cohen's d because it adjusts for biases resulting from small sample sizes (Hedges & Olkin, 1985), which was a factor with the data included in this review. Statistical analyses were run in R Studio (Version 3.2.2) using the metafor package (Viechtbauer, 2010). See Figure 2.6 for a sample of the R syntax and refer to Appendix B for the full syntax.

```
>###Setting the working directory###
>setwd("Z:/1 DPhil/7 MMSR Study/R_Working_Directory")
>
>###Teacher Negative Strategies###
>library(metafor)
>TNS<-read.csv("TNS.csv", header=T)
>TNS
>MLM_TNS<-rma.mv(yi=hedges_g, V=V_g, data=TNS, random=list(~1|esid, ~1|studyid),
slab=studyoutcome)
>summary(MLM_TNS)
>l2_TNS_es<-
(MLM_TNS$sigma2[1])/(MLM_TNS$sigma[1]+MLM_TNS$sigma[2]+mean(TNS$V_g))
>l2_TNS_study<-
(MLM_TNS$sigma2[2])/(MLM_TNS$sigma[1]+MLM_TNS$sigma[2]+mean(TNS$V_g))
>l2_TNS_es
>l2_TNS_study
>
```

Figure 2.6 Excerpt of R syntax

Sample size considerations. Beyond summarising the effectiveness research in a given area and estimating a grand mean effect size, meta-analysis is often conducted to increase statistical power (Cohn & Becker, 2003). This use raises the important question

of whether there exists a minimum sample size below which meta-analysis (multilevel or otherwise) should not be conducted. Theoretically, one could conduct a meta-analysis with two studies. However, generally, it is assumed that the larger the sample size of studies, the greater the statistical power to detect effects with precision. This is indeed the case when computing fixed effects models (Cohn & Becker, 2003). Yet, because random effects models are not estimating one population effect (as in fixed effects models), but rather are estimating the mean and distribution of population effects, it is not always the case that increasing the sample size of studies will result in increased statistical power (Cohn & Becker, 2003). This is due to the possibility that additional studies might increase the group level (between-study) variance component, thus increasing the standard error and reducing statistical power (Cohn & Becker, 2003).

Despite the logic behind a random effects model and the importance of the between-study variance, the question of sufficient sample size persists. Maas and Hox (2005) discuss how, the constraint around sample size is a greater issue at the higher level (i.e., study or group-level). They conduct a simulation study to examine different group and total sample sizes and their effects on regression coefficients, variance components, and standard errors. They found that when the group-level (e.g., number of studies) sample size was 10, regression coefficients and individual-level variance components were estimated without bias, but between-study (i.e., group-level) variance components were too big and standard errors too small (Maas & Hox, 2005). As a result, they conclude that a sample size of 10 at the group level is insufficient for multilevel models. However, Valentine, Pigott, and Rothstein (2010) explain that the lower the between-group heterogeneity is in a random effects model, the greater the estimated power. While they would still recommend having higher sample sizes at the group level (i.e., more included studies in the analysis), they argue that there is still reason to conduct the analysis even

when power is estimated to be low, because an ‘additional benefit arising from doing the meta-analysis is that it allows the reviewers to compute and interpret confidence intervals and as such adds information beyond what is revealed in the individual studies’ (Valentine et al., 2010, p. 239). They also maintain that conducting the analysis even in cases of low power offers the potential for a more valid conclusion than simple vote counts or other strategies.

In this review and multilevel meta-analysis, which anticipated a small number of included studies (group-level), grand mean effect sizes were calculated with the understanding that the small number of studies could limit statistical power and lead to Type II Error in accepting the null hypothesis when it should really be rejected. Therefore, following the above-discussed guidance, the focus of the results will be more on the magnitude and direction of effect in indicating trends rather than solely focusing on tests of statistical significance.

Dealing with missing data. In cases of incomplete quantitative data, study authors were contacted to obtain additional information. Failing clarification, studies were excluded from the meta-analysis but not the narrative synthesis if post-intervention data (e.g., n-values for experimental/control groups or standard deviations) were not sufficiently reported. This strategy maintains the integrity of the meta-analysis, since assuming zero attrition at post-intervention can distort effect sizes, and means reported without standard deviations do not accurately describe the data set.

Assessment of heterogeneity. As the quantitative strand of this review included only studies implementing an RCT design to evaluate IY TCM against a treatment-as-usual control with a specified age-range and setting, clinical heterogeneity was not anticipated. However, due to the variety of locations implementing IY TCM and the potential for substantial variation of ‘treatment-as-usual’ in each location, tests were

conducted to assess statistical heterogeneity using the I^2 statistic as a signal-to-noise ratio illustrating the percentage of across-study variation due to underlying differences of the studies as opposed to chance (Higgins, Thompson, Deeks, & Altman, 2003). Additionally, the Q test (χ^2 test) served as a second assessment of heterogeneity, even though it is sensitive to the number of included studies, unlike the I^2 statistic (Borenstein, Hedges, Higgins, & Rothstein, 2009).

Assessment of reporting biases. The Cochrane Handbook recommends drawing funnel plots to test for publication bias only when working with ten or more included studies. As this minimum criterion was not met, no funnel plot was drawn for this review.

Meta-analysis of risk subgroups. Studies included in this review took different approaches when examining subgroups of children. This review *a priori* planned to conduct multilevel meta-analyses of risk subgroups in the areas of child's baseline at-risk level, child sex, and children identified with special educational needs. The only risk subgroup reported in any of the included studies was the baseline level of risk for behavioural difficulties; therefore this is the only risk subgroup meta-analysed.

Sensitivity analysis. In the event of determining any of the included quantitative studies was at high risk of bias, sensitivity analyses would have been run to isolate these studies from those at a lower risk of bias. Because none of the quantitative studies was deemed at high risk of bias, sensitivity analysis was not run. After identifying the included studies and extracting the data, it was discovered that three of the four studies included in the multilevel meta-analysis had collected child-level data on a cross-section of children from the classroom to sample low, medium, and high-risk children. The remaining one study only collected data from those children who were rated by their teacher at baseline as having high risk of conduct problems. Therefore, to explore the influence of this risk subgroup on the overall grand mean, *post hoc* sensitivity analyses were run on all of the

child outcomes by excluding the data from this one study. This will be discussed further in the results section.

Data Extraction and Analysis – Qualitative Strand

Extraction. Qualitative data was extracted into NVivo 11 for analysis.

Quality assessment. Although it is accepted practice to conduct a risk of bias assessment on quantitative studies to assess the quality of evidence in a review, consideration of the quality of individual qualitative studies raises unique questions. These questions arise because the practice of critically appraising qualitative research has reached no consensus on how it should be accomplished (Britten et al., 2002; Chapple & Rogers, 1998). Should qualitative studies be excluded based on quality, and how is quality appraised (Walsh & Downe, 2005)? Some argue that the inability to define a ‘unified’ qualitative research paradigm means that each study should be judged on its own merits, rather than on one set of guidelines (Rolfe, 2006). Conversely, Dixon-Woods, Shaw, Agarwal, and Smith (2004) posit that certain characteristics of qualitative research reporting and conduct are constant, regardless of methodology. Sandelowski and Barroso (2007) recommend appraisal, while advising against excluding studies based on perceived quality. Thus, including or excluding studies on the basis of quality is understandably contentious, although assessment of quality can become a focus for analysis.

In the current study, qualitative studies were assessed for quality using an adapted EPPI-Centre checklist, which has been used in other qualitative meta-syntheses (Bonell et al., 2013). This checklist assesses studies on rigour in sampling, data collection, data analysis, whether findings were grounded in the data, and the breadth/depth of the findings. It also considers reliability, trustworthiness, and usefulness of the study to the review. Studies were not excluded on the basis of a poor quality assessment; however,

such ‘low quality’ studies did not serve as the initial point for reciprocal translation in the meta-synthesis process.

Dealing with missing data. Where findings from qualitative data were sparsely reported, further elaboration from authors was not sought, because the qualitative meta-synthesis is based on the authors’ interpretations/findings, and to ask for specific *post hoc* elaboration on specific areas could have changed the perspectives of these findings.

Qualitative meta-synthesis. Analysis for this qualitative meta-synthesis was based in a grounded theory approach, which operates both inductively and deductively, and aims to generate, not test, theory (Glaser & Strauss, 1967). When coding qualitative data, the analyst transforms the data from its original state to a conceptual level (Corbin & Strauss, 2008). To start, open codes are assigned to the data, with the goal of identifying the many various inquiries one could pursue with the data (J. Green & Thorogood, 2014). Then, the analyst connects the categories by establishing their relationships to one another through axial coding (J. Green & Thorogood, 2014). The analysis proceeds towards progressively more abstract and theoretically informed codes (hence the deductive parts of the approach) with the emergence of core categories that remain substantiated in the original data (J. Green & Thorogood, 2014). Throughout this process, the analyst is continuously making comparisons within and between incidents from the data. This method of constant comparison includes four stages, as outlined by Glaser and Strauss (1967): ‘(1) comparing incidents applicable to each category, (2) integrating categories and their properties, (3) delimiting the theory, and (4) writing the theory’ (p. 105). Following this approach will help explicate patterns between the data and identify any negative case, which does not fit into emerging patterns (Corbin & Strauss, 2008). Any such examples sitting outside of the emerging theory will be considered from other perspectives, allowing for deeper understanding of the variety of experiences and views among people.

Other types of qualitative meta-synthesis methods, including thematic analysis, were considered for this analysis, but were rejected as they were not the best fit for the research question driving this project. A more detailed description of the various approaches to qualitative meta-synthesis and the many issues researchers need to consider when conducting a qualitative meta-synthesis has been published (Nye, Melendez-Torres, & Bonell, 2016b) and can be found in Appendix C. This article is accompanied by a Context and Implications document (Nye, Melendez-Torres, & Bonell, 2016a).

Sensitivity analysis. Sensitivity analyses were not applied to the qualitative strand, as specific methods for doing so are as yet unclear and the topic itself is contentious within the discourse about qualitative meta-synthesis (Nye et al., 2016b).

Cross-synthesis

After the quantitative multilevel meta-analysis and qualitative meta-synthesis were complete, the results were cross-synthesised to examine common and divergent themes and develop new implications for research and practice. It also allows for weighting of the evidence, in considering the scope of any overlapping themes between the quantitative and qualitative themes. In this way, the synthesising argument addresses the direction and magnitude of evidence and sets an agenda for future research to either fill in gaps or to strengthen the weight of evidence in areas with low coverage. To this end, framework analysis was applied to the findings of the meta-analysis and meta-synthesis. Framework analysis was originally created by the National Centre for Social Research to develop findings relevant for practice and policy (J. Green & Thorogood, 2014). Its four-step process begins with familiarisation of the data, proceeds to thematic analysis, followed by indexing (applying codes), and finishes with charting, or arranging data by theme (J. Green & Thorogood, 2014). The third step (indexing) and fourth step (charting) are particularly

appropriate for large datasets, as this method was originally developed to handle large amounts of data for policy and practice.

While the amount of data included in this review was not so large as to require the structured framework approach, the method was nevertheless still appropriate and applicable in that it provided the necessary organisation and transparency to ensure that the main results and findings from both the quantitative and qualitative syntheses were kept in cross-synthesis. Adapting the integrative grid from Flemming (2009), data for the cross-synthesis were charted in two ways. First, key results of the quantitative synthesis were inserted into column headings and the qualitative studies were inserted into the rows. Findings from the qualitative studies populated the cells, indicating the intersection between the quantitative outcomes and qualitative studies. Second, key themes from the qualitative synthesis were inserted into the column headings and the quantitative studies were put in each row. Results from the multilevel meta-analysis and narrative synthesis occupied the cells to illustrate the joining of qualitative findings and quantitative studies. By creating these integrative grids both ways, neither the effectiveness data nor the experiential data were valued over the other as the 'starting point' and discussion could subsequently emerge from both explanatory and exploratory positions. Visually, these grids also provide an illustration of which themes/concepts were commonly addressed between the quantitative and qualitative research (regardless of agreement or disagreement), which themes/concepts were only addressed by one of the two strands of research, and which studies covered a broad or narrow range of outcomes/topics.

Chapter 3 : Effectiveness of a teacher classroom management programme – Results of the quantitative systematic review and multilevel meta-analysis

Results

In this review, references are made to included studies and records. ‘Records’ refer to the individual papers (e.g., journal articles, reports, dissertations, conference abstracts) retrieved. ‘Studies’ refer to the research projects conducted on the IY TCM programme. Because of the overarching nature of the studies, many of them encompass multiple records. This is due to various results from research projects (studies) being written up for dissemination in different formats (records).

Description of studies

Results of the search. After conducting the electronic database and website searches and contacting relevant researchers in this field, all retrieved records were imported into EndNote X7 and duplicates removed (see Figure 3.1 for PRISMA flowchart). Two researchers independently screened all titles and abstracts for the 3,875 remaining records. Based on this initial screening, 3,812 records were deemed irrelevant to the review, leaving a potential 63 records to be assessed for inclusion based on their full text. Two of these potential studies were doctoral dissertations that could not be retrieved in full-text format (see ‘Studies awaiting classification’ section). The same two researchers independently read the 61 full-text reports retrieved and judged each record to be included or excluded based on the pre-specified criteria. These independent decisions were then jointly compared, and the ultimate decision was made by consensus. There were no disagreements, and consensus was reached on all records. Three records (Ford et al., 2012; Hansford et al., 2015; D. W. Murray, 2009) were protocols or grants describing on-going studies. Forty-four records were excluded for not meeting inclusion criteria. The remaining 14 records were included in this review and jointly described nine studies that

fit into the qualitative, quantitative, or both sections of this paper. Of the five quantitative studies included in this review, only four were included in the multilevel meta-analysis due to insufficiently reported outcome data in one study.

Included studies. Nine studies reported in 14 papers met the inclusion criteria for this mixed methods review: four studies in six records were classified as quantitative; four studies in four records were classified as qualitative; and one study in four records was classified as both qualitative and quantitative. The nine included studies took place across five different countries: one study in England and two studies each in Ireland, Jamaica, the United States, and Wales. This section highlights the characteristics of all five included studies in the quantitative strand (see Tables 3.1 and 3.2) to provide context for the subsequent findings from the multilevel meta-analysis.

Limerick, Ireland Study. This mixed methods study combined the experimental design of a cluster randomised controlled trial across 11 schools with 22 teachers and 217 index children and a two-part qualitative process evaluation with class teachers. This study's findings are detailed across four papers, including two reports (McGilloway et al., 2010; McGilloway et al., 2012), one PhD dissertation (Hyland, 2014), and one peer-reviewed journal article (Hickey et al., 2015). Only the first report (McGilloway et al., 2010) and the peer-reviewed journal article (Hickey et al., 2015) presented findings included in the quantitative strand of this review.

The randomised controlled trial took place in the south-west of Ireland and included mostly urban schools ($n = 8$), with few semi-urban schools ($n = 2$), and a rural school ($n = 1$). Two teachers from each school were included in the study, with one teacher receiving the IY TCM intervention and the other teacher serving as the waitlist control (teachers received training the following year). At baseline (and prior to

randomisation into intervention or control groups), all participating teachers completed the

teacher version of the Strengths and Difficulties Questionnaire (Goodman, 1997). Based on these scores, a subset of 12 index children were invited to participate in the study, representing approximately four children with low-risk of behavioural problems, four with medium-risk, and four with high-risk. Researchers excluded children from the study for three reasons: if there was no parental consent, if the child was absent from school more than once during data collection, and if the child was identified as having a developmental disorder. To reduce potential bias, an independent statistician randomly allocated teachers to study condition using a computer-generated random number sequence.

Teachers in the intervention group attended a full day course once a month for five consecutive months. In between each training session, the IY TCM group facilitators called the intervention teachers once to discuss implementation of learned strategies. Additionally, intervention teachers kept a written diary and were able to discuss their progress within the intervention group at each training session. This study reported outcomes for both teachers and children. At post-intervention, teachers completed self-report measures on their classroom management strategies as well as on how useful they thought the IY TCM strategies were for their classrooms. They also completed reports on child problem behaviour and child well-being for the index children in their class. Members of the research team visited the classrooms to conduct independent observations of both the teacher and the index children to measure teacher management strategies and child behaviour. Analyses of child-level outcomes were adjusted using intraclass correlation coefficients (ICCs).

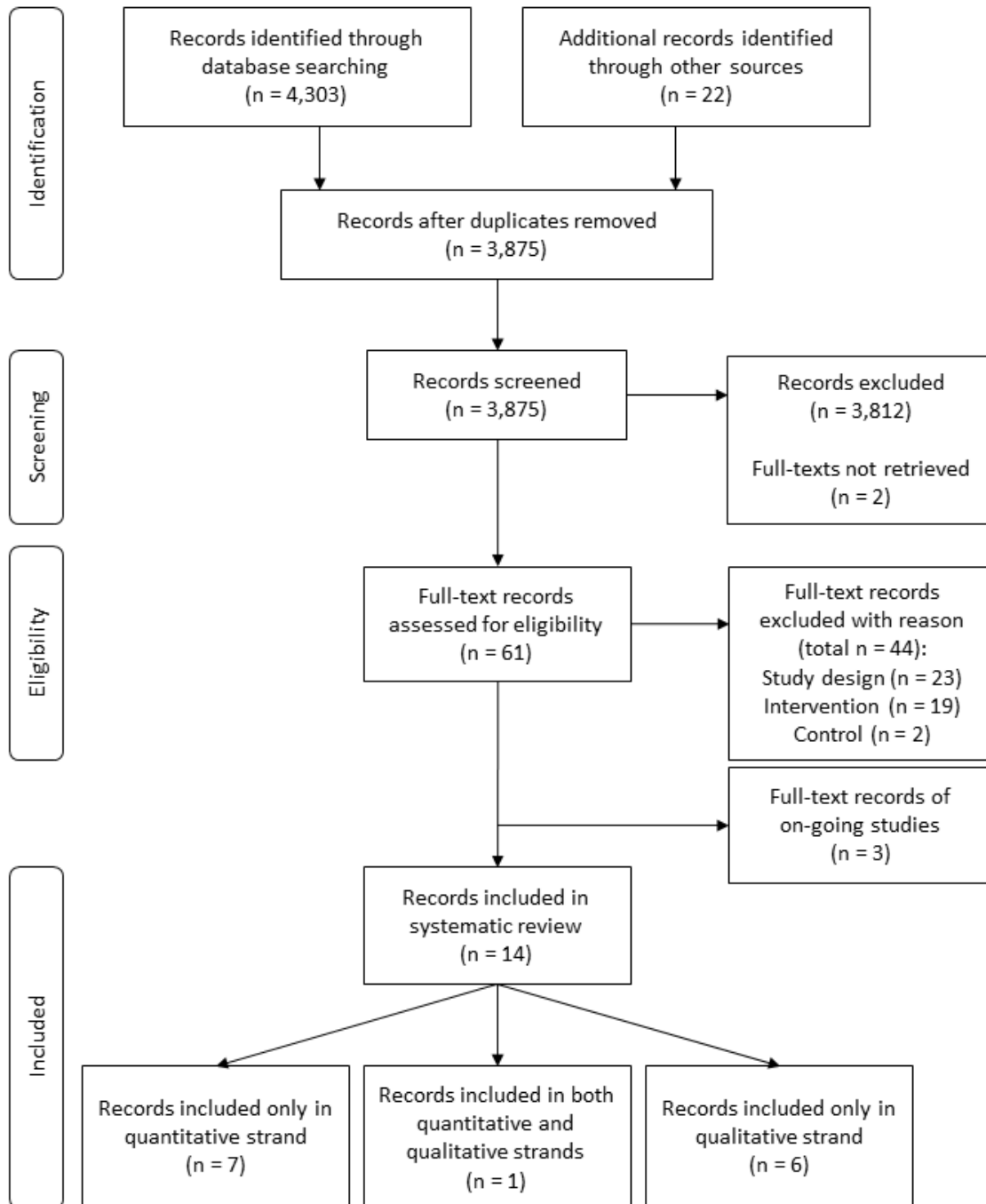


Figure 3.1 PRISMA flowchart

Table 3.1 Included studies

Included studies for Quantitative Strand	
Study ID	Records
Limerick, Ireland	McGilloway et al., 2011
	Hickey et al., 2015
Kingston, Jamaica (2)	Baker-Henningham et al., 2012
Rhode Island, USA	Shepard et al., 2008
North Carolina, USA	Murray et al., 2012
	Murray et al., 2014
North Wales (2)	Martin, 2009
	Hutchings et al., 2013

Kingston, Jamaica (2) Study. This cluster randomised controlled trial is reported in one peer-reviewed journal article (Baker-Henningham et al., 2012). Twenty-four urban pre-schools participated in the study, with a combined 73 classrooms and 225 index children. In this study, the school was the unit of randomisation, so all teachers in intervention schools received the IY TCM training. Prior to randomisation, teachers completed a 10-item questionnaire to rate the children in their classrooms on level of conduct disorder. This baseline served to identify three children from each classroom with high-risk for behavioural problems. Researchers excluded children from the study for three reasons: if the child lived in an institution or had a sibling who was enrolled in the study, if the child's attendance was less than 70 per cent, and if the child was identified as having a developmental disability. An independent statistician carried out the computer-generated randomisation to allocate each participating school to either the intervention or control group.

The IY TCM programme was culturally adapted based on qualitative focus groups, a pilot study, and in-depth qualitative interviews post-pilot study (Baker-Henningham, 2011; Baker-Henningham, Walker, Powell, & Gardner, 2009a, 2009b), with training

occurring over eight days rather than five or six days, and with practical activities tailored to reflect a Jamaican classroom rather than the original American classrooms. In addition to the training, teachers received one hour of in-class coaching each month for four months. Teachers in the control group schools attended training-as-usual in-service workshops during the course of the trial. This study reported only child-level outcomes, with the primary outcome at post-intervention being researcher observations of child behaviour in the classroom. Secondary outcomes at post-intervention included both teacher- and parent-reported child behaviour and child attendance. Analysis of outcome data accounted for the nested nature of the children within classrooms/schools with the intraclass correlation coefficients.

Rhode Island, USA Study. This cluster randomised controlled trial is reported in one conference poster (Shepard, Doctoroff, Dickstein, & Seifer, 2008). Thirteen class teachers along with 13 teaching assistants participated in this study, which was part of a larger partnership between the affiliated university and Head Start site (urban). In total, 213 children participated in the study. Specific details relating to the process of randomisation are not reported, but seven teaching pairs (class teacher and teaching assistant) and one education manager were randomly assigned to the intervention arm of the study. Baseline data was collected one month prior to the start of the intervention. The IY TCM programme took place over the course of six full-day workshops. Teaching pairs in the control group received standard Head Start training. Data were collected on classroom environment, teacher classroom management strategies, teacher-pupil relationships, child social skills, and child behaviour problems. This conference poster reported outcomes across all of these domains, but the absence of fully-reported statistical outcomes (e.g., intervention versus control *n*-values, standard deviations) prevented further use of these data in the multilevel meta-analysis.

North Carolina, USA Study. This cluster randomised controlled trial is reported in two conference abstracts (D. W. Murray, Murr, & Rabiner, 2012; D. W. Murray, Rabiner, & Carrig, 2014) as well as in the grant's structured abstract (D. W. Murray, 2009). The first conference abstract (D. W. Murray et al., 2012) describes the study as including a total of 44 class teachers from six participating rural schools. In this study, a block design was used to randomise grade levels within schools, such that each school had a minimum of one grade level assigned to the intervention group. The researchers chose to randomise at grade level (rather than teacher or school levels) because they had few concerns about contamination across grade levels within school. The IY TCM programme was delivered over the course of five full-day workshops. In between the monthly training sessions, participating teachers were able to meet with their IY TCM group facilitators for individual consultation on implementing the learned strategies in their classrooms. This first conference abstract only reports on preliminary teacher/classroom outcomes, including emotional support, classroom organisation, and instructional support as well as on competent and harsh management strategies.

The second conference abstract (D. W. Murray et al., 2014) includes both teacher and pupil data for 97 class teachers (47 assigned to intervention condition with 45 actually participating) and 1,276 children (data analysis conducted on 598 children in intervention classrooms and 560 children in control classrooms; total 1,158 children). The same blocked randomisation and training schedule as described in the first conference abstract still apply. Baseline data collection took place at the start of the academic year (subsequent to teacher consent and randomisation) and involved researcher observations (blind to condition) and teacher-reports of child behaviour. Data reported in this conference abstract include classroom climate, competent management strategies, child emotional regulation, child prosocial behaviour, and child inattention. The researchers

applied multilevel modelling statistical analyses to account for the clustered nature of the data.

According to the structured abstract from the grant summary, teachers in the control condition were placed on a waiting list to receive the IY TCM training during the following academic year. During their waiting year, these teachers attended their normal professional development programme as provided by their school district.

North Wales (2) Study. This cluster randomised controlled trial is described in two reports: one PhD dissertation (Martin, 2009) and one peer-reviewed journal article (Hutchings et al., 2013). Eleven schools (12 classrooms) with Welsh as the main language of instruction participated in this study. Eight of the classrooms had a full-time teacher who participated, and the remaining four classrooms had two part-time teachers job sharing. In these classrooms, the part-time teacher who was in the classroom for more instruction time during the week participated in the study. As a result of teachers' baseline assessments of children using the teacher Strengths and Difficulties Questionnaire (Goodman, 1997), 107 index children participated in the study, with most ($n = 80$) falling into the low-risk group and few ($n = 27$) falling into the high-risk group for behaviour problems. Following baseline measures, an independent researcher carried out the randomisation process using a random number generator. Classrooms were first paired based on the size of both the school and the classroom as well as whether it was urban/rural. One classroom from each pair was randomly assigned to the intervention arm of the study and the other was subsequently assigned to the control arm. Teachers in the intervention group attended five full-day IY TCM training sessions, each scheduled a month apart. Teachers in the control arm received no in-service trainings during the course of the study, but as this was a waiting list control they received the IY TCM training the following year. The study reported both child and

Table 3.2 Characteristics of included studies

Study ID	Participants	Intervention	Control	Outcomes	Analysis
Limerick, Ireland	11 primary schools (8 urban, 2 semi-urban, 1 rural) 22 teachers (2 per school) 12 children per class	IY TCM (5 sessions)	Wait-list control	Teacher outcomes Self-reported use of positive and negative strategies (frequency) and perceived usefulness of these strategies Observed positive strategies, praise, negative strategies, indirect commands, direct commands, and no opportunity for child to respond Child outcomes Teacher-reported child behaviour difficulties and prosocial behaviours Observed negative behaviour, positive behaviour, compliance, and non-compliance	Teacher outcomes ANCOVA Child outcomes Robust multiple linear regression
Kingston, Jamaica (2)	24 community pre-schools (urban) 3 children per class (high-risk only)	IY TCM (8 sessions; Culturally adapted to Jamaican context from pilot study in Baker-Henningham et al., 2012)	Treatment-as-usual (Regular, Ministry of Education in-service workshops; Received identical educational materials)	Child outcomes Observed child behaviour Teacher-reported child behaviour Parent-reported child behaviour Child attendance Parent outcomes Self-reported attitude to school	Child outcomes Multilevel regression models for continuous data, Random-effects logistical regression models estimated with Markov chain Monte-Carlo methods for binary data
Rhode Island, USA	13 lead teachers 13 teacher assistants 231 children	IY TCM (6 sessions)	Treatment-as-usual (Standard Head Start training)	Teacher outcomes Teacher-reported positive and negative strategies, approach with parents, and relationship with children Classroom outcomes Observed classroom environment Child outcomes Teacher-reported child social skills and child behaviour problems	Not addressed on poster
North Carolina, USA	44 teachers (2012) and 97 teachers (2014) 1276 children	IY TCM (5 sessions)	Wait-list control	Teacher outcomes Observed teacher management strategies Classroom outcomes Observed classroom environment Child outcomes Teacher-reported child social-emotional outcomes, child behavioural and attention outcomes, and child academic outcomes in literacy/reading and mathematics	Repeated measure GLM to evaluate between group differences, Multilevel modelling to account for nested data
North Wales (2)	12 teachers (8 full-time, 4 part-time) 107 children Welsh-speaking classrooms	IY TCM (5 sessions)	Wait-list control (No in-service or other professional development training sessions offered during trial)	Teacher outcomes Observed teacher management strategies Teacher-reported satisfaction with programme Child outcomes Observed child behaviours	Hierarchical linear modelling (HLM)

teacher outcomes, and with the exception of the teachers' baseline reports of child behaviour, all measures collected were observations of child and teacher behaviour conducted by members of the research team blinded to study condition to minimise bias. Analysis of child-level outcomes appropriately included intracluster correlation coefficients to account for the nested nature of the child-level data within their classroom (i.e., teacher-level) environment.

Excluded studies. Forty-two records were examined and excluded from the quantitative strand of this review for failing to meet all of the *a priori* inclusion criteria set forth (see Table 3.3). Most records were excluded due to study design (i.e., not RCTs) or intervention (i.e., not IY TCM).

On-going studies. Three studies were classified as on-going for the quantitative strand of this review, although of those three, one was included in the qualitative strand (Devon) and one had preliminary data included in the quantitative strand (North Carolina).

Devon, England Study. The mixed methods study in England, which is included in the qualitative strand of this review, is cross-listed as an on-going study for the quantitative strand because its quantitative outcomes are yet to be completed. This study's main protocol (Ford et al., 2012) and process evaluation protocol (Hansford et al., 2015) describe a large, cluster-randomised controlled trial occurring over three cohorts of participating teachers/schools and a two-year follow-up period. The last round of follow-up quantitative data will be collected in early 2017, and it is expected that results will be disseminated shortly afterwards.

North Carolina, USA Study. The abovementioned study in North Carolina, USA, whose two conference abstracts are included as preliminary data in the quantitative strand of this review, has concluded, but the final outcome results and reports are currently in the process of being peer-reviewed and published. As such, the main study's data could not

yet be included in this review. It is expected that the full journal article(s) detailing the complete post-intervention findings will provide additional details not only on the full results but also on the methods and analyses applied for an updated risk of bias assessment.

Missouri, USA Study. According to the grant description (Reinke, Herman, & Stormont, 2010), this efficacy study includes 100 teachers across 20 schools in an urban school district in the Midwest region of the United States. Quantitative data was collected on teacher behaviours, child behaviours, and child academic achievement. Although this study is now concluded, its final reports and results have yet to be published and therefore are not included in this review. However, related publications from this study can be found in the excluded studies section, as these reports focused on various aspects of the study's implementation and preliminary or sub-component data, which did not meet inclusion criteria for this review (Reinke, Herman, Stormont, Newcomer, & David, 2013; Reinke, Stormont, Herman, & Newcomer, 2014; Reinke, Stormont, Herman, Wang, et al., 2014; Reinke, Stormont, Webster-Stratton, Newcomer, & Herman, 2012).

Studies awaiting classification. Two reports (both dissertations) could not be retrieved in their full-text format during the course of this review. No peer-review journal articles, conference abstracts, or other types of publications were found that related to either of these dissertations to supplement the information provided in the abstracts of these reports to make a firm decision. The Kennedy (2004) and Wetherall (2014) dissertations are therefore unable to be classified as either included or excluded at this time.

Table 3.3 Excluded studies

Record	Reason for exclusion
Considered only for quantitative strand	
Baker-Henningham, Walker, Powell, & Gardner, 2009a	Did not isolate IY TCM
Baker-Henningham, Walker, Powell, & Gardner, 2009b	Did not isolate IY TCM
Carlson, Tired, Bender, & Benson, 2011	Not an RCT
Dicke, Elling, Schmeck, & Leutner, 2015	Not about IY TCM
Fergusson, Horwood, & Stanley, 2013	Not an RCT
Fernandez, Gold, Hirsch, & Miller, 2015	Not about IY TCM
Foster, Johnson-Shelton, & Taylor, 2007	Not an RCT
Foster, Olchowski, & Webster-Stratton, 2007	Not an RCT
Fox, Hemmeter, Snyder, Binder, & Clarke, 2011	Not about IY TCM
Gaspar (2013)	Not an RCT
Herman, Borden, Reinke, & Webster-Stratton, 2011	Did not isolate IY TCM
Hobbel & Drugli, 2013	Not about IY TCM
Hutchings, Bywater, Daley, & Lane, 2007	Not about IY TCM
Malti, Ribeaud, & Eisner, 2011	Not about IY TCM
Martin, Daley, Hutchings, Jones, et al., 2010	Not an RCT
Olchowski, Foster, & Webster-Stratton, 2007	Not an RCT
Rappuhn, 2014	Not appropriate control condition
Raver, Jones, et al., 2008	Did not isolate IY TCM
Reid & Webster-Stratton, 2001	Did not isolate IY TCM
Reid, Webster-Stratton, & Hammond, 2003	Did not isolate IY TCM
Reinke, Herman, Stormont, Newcomer, et al. 2013	Not an RCT
Reinke, Stormont, Herman, & Newcomer, 2014	Not an RCT
Reinke, Stormont, Herman, Wang, et al., 2014	Not about IY TCM
Reinke, Stormont, Webster-Stratton, et al., 2012	Literature review
Sherhoff & Kratochwill, 2007	Not appropriate control condition
Snyder, Low, Schultz, Barner, Moreno, et al., 2011	Adapted IY TCM
Webster-Stratton, Reid, & Hammond, 2001	Did not isolate IY TCM
Webster-Stratton, Reid, & Hammond, 2004	Did not isolate IY TCM
Webster-Stratton, Reid, & Stoolmiller, 2008	Did not isolate IY TCM
What Works Clearinghouse, 2011	Literature review
Whear, Thompson-Coon, Boddy, Ford et al., 2013	Systematic review
Winther, Carlsson, & Vance, 2014	Not about IY TCM
Considered both for quantitative and qualitative strands	
Colletta, Latimore, Reinke, & Herman, 2013	Informational study brochure
Hutchings, 2012	Programme report
Pidano & Allen, 2015	Literature review
Webster-Stratton, 2001	Literature review
Webster-Stratton & Bywater, 2014	Literature review
Webster-Stratton & Herman, 2010	Programme development and fidelity
Webster-Stratton & Reid, 2007	Literature review

Risk of bias in included studies

It is recognised that the type of publication/report influences the available space for study authors to detail aspects about their trial methods. A PhD dissertation should have considerably more details than one could expect to find in a conference abstract. Given these realities and as previously detailed in the methods section (Chapter 2), studies were rated as having an unclear risk of bias where incomplete reporting resulted in a lack of information to make a risk of bias judgement. Excerpts from the relevant records are provided in Appendix D to support the risk of bias ratings applied to each study and described below. A summary of the risk of bias of included studies can be seen in Figure 3.2. The interrater percentage agreement was greater than 85 per cent across all questions on each study.

Recruitment bias. To avoid bias during recruitment, clusters and individuals should be recruited prior to random assignment to either the intervention or the control group. Four of the five studies clearly specified that such a process had occurred and were rated as having a low risk of recruitment bias: Limerick, Kingston (2), North Carolina, and North Wales. One study, Rhode Island, did not clearly report this aspect of the study design and was therefore rated as of unclear risk.

	Recruitment bias	Selection bias		Performance bias	Detection bias	Attrition bias		Statistical analysis bias		Reporting bias	
	Individuals recruited after clusters assigned to intervention/control	Random sequence generation	Allocation concealment	Blinding of participants and personnel	Blinding of outcome assessment	Incomplete outcome data – cluster level	Incomplete outcome data – individual level	Incorrect analysis given clustering of data	Intention-to-treat (ITT) analysis	Reporting of baseline comparability of clusters	Selective reporting
Limerick, Ireland	Low risk	Low risk	Low risk	Unclear risk	Unclear risk	Low risk	Low risk	Low risk	Unclear risk	Low risk	Low risk
Kingston, Jamaica (2)	Low risk	Low risk	Low risk	Unclear risk	Unclear risk	Low risk	Low risk	Low risk	Low risk	Low risk	Unclear risk
Rhode Island, USA	Unclear risk	Unclear risk	Unclear risk	Unclear risk	Unclear risk	Low risk	Unclear risk	Unclear risk	Unclear risk	Low risk	Unclear risk
North Carolina, USA	Low risk	Unclear risk	Unclear risk	Unclear risk	Unclear risk	Low risk	Unclear risk	Unclear risk	Unclear risk	Low risk	Unclear risk
North Wales (2)	Low risk	Low risk	Low risk	Unclear risk	Low risk	Low risk	Low risk	Low risk	Low risk	Low risk	Low risk

Figure 3.2 Risk of bias assessment

Selection bias. It is important for randomised controlled trials to have a truly random sequence generation to avoid inserting bias into the selection process. Three of the five included studies specified appropriate methods for generating a random sequence and are thus considered to have a low risk of bias at this stage: Limerick, Kingston (2), and North Wales (2). The remaining two included studies did not specify enough details to make a determination on the likelihood of risk entering the study at this stage, and were rated as having an unclear risk: Rhode Island and North Carolina.

Steps should also be taken to ensure that bias is not introduced between random sequence generation and communicating assignments to participants (allocation concealment). The same three studies with sufficient detail on sequence generation provided information in their reports to indicate a low risk of bias at this stage of the process: Limerick, Kingston (2), and North Wales (2). Both of the Rhode Island and North Carolina studies were rated as at unclear risk of bias at this stage due to lack of reporting on this topic.

Performance bias. As with so many RCTs of social programmes (as opposed to drug trials), double-blind studies were not feasible. In any study examining the effects of IY TCM, it would be impossible for the teacher participants to be blind to their assignment (i.e., intervention or control). However, to guard against Hawthorne effects, study designs could incorporate an active control group, where the control teachers receive some sort of attention unrelated to classroom management. Barring an active control, the only way to introduce some protection for performance bias is to keep the research team blind to assignments and request that participants not share their assignment with the research staff. Due to the nature of all of these included studies and the absence of any active controls, it is determined that all five included studies are at unclear risk of bias for this stage.

Detection bias. When outcome assessors are aware of intervention or control group status, they might inadvertently introduce detection bias by noticing or looking for signs that confirm expected outcomes. With the inclusion of teacher-reported data (on child outcomes and self-reports on their own outcomes) in many of these studies, blinding of outcome assessment is broken. In four of the five included studies, this teacher-reported data was balanced with independent observational data collected by researchers blind to assigned condition: Limerick, Kingston (2), Rhode Island, and North Carolina. Therefore, these four studies received a classification of unclear risk for this stage. The North Wales (2) study only considered outcome data collected by observers blind to assigned condition. As such, it is rated as at low risk of bias for this stage.

Attrition bias. Bias can also be introduced into studies when clusters and/or participants systematically drop out before the conclusion of the project. Researchers should address cases where participants have left early or were unavailable at post-test. None of the five studies experienced cluster attrition and were all rated as at low risk on this item. The Limerick, Kingston (2) and North Wales (2) studies were rated at low risk for incomplete outcome data bias based on either determining no difference between dropouts and completers or on having no attrition among participants. The Rhode Island, and North Carolina studies were both deemed to have unclear risk at this stage for not clearly or completely addressing attrition in the reports.

Statistical analysis bias. When conducting educational research, it is common for data to be nested at different, naturally occurring levels, such as children (level 1) nested within a teacher's classroom (level 2) nested within a school (level 3). When an intervention is delivered at a level higher than that of data collection, it cannot be assumed that the data are statistically independent due to their clustered nature. Assuming statistical independence in this scenario would lead to an overestimation of treatment

effect. Therefore, adjusting the standard error using the intracluster correlation coefficient (ICC) avoids Type I Error, or concluding that there is an effect when there is not. The Limerick, Kingston (2), and North Wales (2) studies report ICCs or otherwise report analyses that rigorously take into account the nested data. Therefore, these studies are rated as at low risk for bias at this stage. The Rhode Island and North Carolina studies do not list ICCs and are rated as having an unclear risk.

Intention-to-treat (ITT) is when researchers include data from all participants who were initially assigned to each condition in their analyses, regardless of whether participants received the treatment of their assignment. Applying an ITT approach during analysis minimises risk of overestimating effects due to possible systematic differences between people in the same group who receive varying dosages of the intended intervention. The Kingston (2) and North Wales (2) studies both received a low risk classification on this measure. The Limerick, Rhode Island, and North Carolina studies did not clearly address this point and were therefore classified as at unclear risk at this stage.

Reporting bias. When researchers collect data through numerous measures but only report results on some of these data, reporting bias enters a study. Similarly, researchers should report on the baseline comparability of the clusters. All of the five studies addressed baseline comparability in their reports and are rated as having low risk of bias on this item. Lack of reporting some results might be due to non-significant findings or negative results, both of which are still important results for understanding the full picture of research in an area. One way to be transparent about data collected/reported is to publish a trial protocol prior to the study's start. Not all study protocols were available, but based on study reports, Limerick and North Wales were at low risk of bias, while there was not enough information about Kingston or Rhode Island to rate other than unclear.

The North Carolina study was reporting preliminary data, and as such it is understood that only some outcomes were going to be available and so that also received an unclear rating.

Effects of interventions

All of the included studies reported outcomes immediately post-treatment. The Rhode Island study measured outcomes at post-intervention and follow-up, but it is unclear exactly how long following post-intervention the follow-up occurred. Since the data were otherwise incomplete (no standard deviations or n-values for intervention and control group data), neither the post-intervention nor the follow-up data from the Rhode Island study were included in the meta-analyses, although the findings are still reported alongside the meta-analysis. All teacher and student data were continuous. Three studies (Limerick, North Carolina, North Wales (2)) reported data from independent observers on teacher classroom management strategies. Two studies (Limerick, Rhode Island) reported data from teachers' self-reports on teacher classroom management strategies.

Three studies (Limerick, Kingston (2), North Wales (2)) looked at child (negative and positive) behaviours via independent observer ratings, and three studies (Limerick, Kingston (2), Rhode Island) used teacher-reports to evaluate child (negative and positive) behaviours.

Only one study examined children's academic readiness (North Carolina) and these results were only preliminary findings presented in a conference paper. As such and based on communication with the first author of that study, these preliminary results are not reported here in anticipation of the forthcoming final report with the complete analyses on academic readiness (D. Murray, personal communication, May 1, 2014).

Similarly, only one study explored teachers' collaboration with parents (Rhode Island). The authors of the Rhode Island study concluded that the 'amount of change in use of [...] positive approaches with parents did not differ between the two groups [...]

and both returned to baseline scores on positive approaches with parents [at follow-up]' (Shepard et al., 2008). Other outcomes relating to parents included parents' attitudes towards school and parents' reports of child behaviour, both from the Kingston (2) study. There were no noticeable differences between intervention and control groups on parents' attitudes to school, with medians (ranges) reported as 28.0 (5-39) and 27.5 (12-37), respectively (Baker-Henningham et al., 2012). The Kingston (2) study also reported on child attendance, as measured by school records, again with no clear difference between intervention and control: 90(48-100) and 88 (30-100), respectively (Baker-Henningham et al., 2012).

Both the Rhode Island studies reported data on classroom environment. The Rhode Island study concluded that the 'difference between [IY TCM] and comparison classrooms in amount of change [of classroom environment scores] was significant, and a large effect size was achieved (Cohen's $d=2.65$) [...] Although scores in both groups declined somewhat by the follow-up assessment, improvements following [IY TCM] were maintained into the next year but scores in comparison classrooms continued to decline' (Shepard et al., 2008, p. no page). The North Carolina study concluded that with 'teaching experience, school, and grade level as covariates, a significant intervention effect (group x time interactions) was found for Positive Climate on the CLASS [$F(1,32)=8.88$, $p=.005$]' (D. W. Murray et al., 2012, p. 4). These data were not meta-analysed because of lack of data reported around the effect size in the Rhode Island study, leaving only the preliminary finding from the North Carolina study.

Teacher and child outcomes that were reported across multiple studies were meta-analysed using a multilevel random effects model as previously described in the methods chapter. See Table 3.4 for a summary of the primary and secondary multilevel meta-analysis results.

Teacher classroom management strategies. Teacher classroom management strategies were evaluated as either negative or positive by independent observers and by teachers' own self-reports. Three studies reported data on teachers' strategies (Limerick, North Carolina, North Wales (2)) using the Teacher-Pupil Observation Tool (T-POT) (Martin, 2009), the Teacher Strategies Questionnaire (TSQ) (Webster-Stratton, 2005), or the Teacher Coder Impressions Inventory (TCI) (Webster-Stratton, Reid, & Hammond, 2001). The three studies contributed seven effect sizes on negative classroom management strategies for a sample size of 72 teachers (35 intervention, 37 control). There was a small, statistically significant effect of the IY TCM intervention on reducing teacher negative classroom management strategies ($g = -0.32, [-0.62, -0.01]$, see Figure 3.3). These studies also contributed seven effect sizes on positive classroom management strategies for the same sample of 72 teachers. There was a moderate, statistically significant effect of the IY TCM intervention on increasing teacher positive classroom management strategies ($g = 0.44, [0.13, 0.74]$, see Figure 3.4).

Table 3.4 Multilevel meta-analysis results for primary and secondary outcomes

Outcome	N(k)	β_{00}	SE	95% CI	I ²	Tau ²	Chi ² (p-value)
Primary							
Teacher negative strategies	7(3)	-0.32*	0.15	(-0.62, -0.01)	L1: 0.00 L2: 0.00	L1: 0.00 L2: 0.00	4.09 (0.66)
Teacher positive strategies	7(3)	0.44**	0.16	(0.13, 0.74)	L1: 0.00 L2: 0.00	L1: 0.00 L2: 0.00	5.16 (0.52)
Child conduct problems	11(3)	-0.20*	0.09	(-0.38, -0.01)	L1: 0.00 L2: 0.00	L1: 0.00 L2: 0.02	10.62 (0.39)
Sensitivity	8(2)	-0.11	0.10	(-0.32, 0.09)	L1: 0.00 L2: 0.00	L1: 0.00 L2: 0.01	5.95 (0.55)
High risk subgroup	8(3)	-0.33***	0.07	(-0.47, -0.18)	L1: 0.00 L2: 0.00	L1: 0.00 L2: 0.00	4.55 (0.71)
Secondary							
Child prosocial skills	6(4)	0.18	0.11	(-0.04, 0.40)	L1: 0.00 L2: 0.00	L1: 0.05 L2: 0.00	13.77 (0.02)
Sensitivity	4(3)	0.13*	0.06	(0.02, 0.24)	L1: 0.00 L2: 0.00	L1: 0.00 L2: 0.00	3.43 (0.33)
High risk subgroup	4(3)	0.31	0.17	(-0.03, 0.65)	L1: 0.00 L2: 0.00	L1: 0.06 L2: 0.00	8.23 (0.04)

N=number of effect sizes (Level 1); k=number of studies (Level 2); β_{00} =Hedges' g ('Grand Mean');

*p<0.05, **p<0.01, ***p<0.001

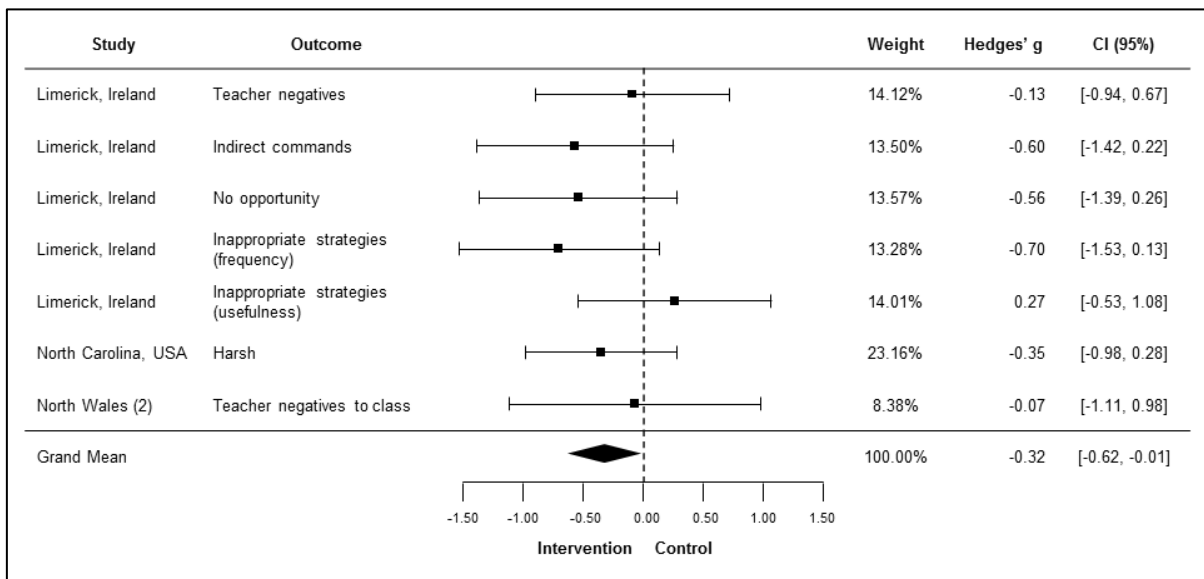


Figure 3.3 Multilevel meta-analysis of teacher negative classroom management strategies

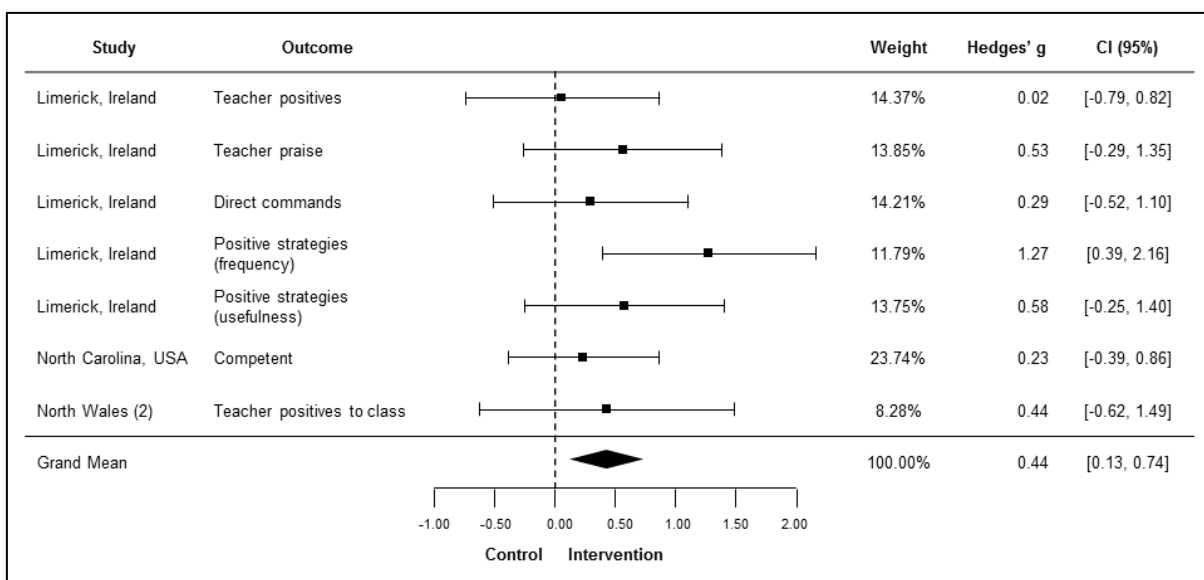


Figure 3.4 Multilevel meta-analysis of teacher positive classroom management strategies

Child conduct problems. Child conduct problems were evaluated by independent observers, teachers, and parents. Three studies reported data on teachers' strategies (Limerick, Kingston (2), North Wales (2)) using the T-POT, the Strengths and Difficulties Questionnaire (SDQ) (Goodman, 1997), or the Dyadic Parent-Child Interaction Coding System (DPICS) (Eyberg & Robinson, 1981) and Multi-Option Observation System for

Experimental Studies (MOOSEs) (Tapp, Wehby, & Ellis, 1995). The three studies contributed eleven effect sizes on child conduct problems for a sample size of 549 children (276 intervention, 273 control). There was a small, statistically significant effect of the IY TCM intervention on reducing child conduct problems ($g = -0.20$, $[-0.38, -0.01]$, see Figure 3.5).

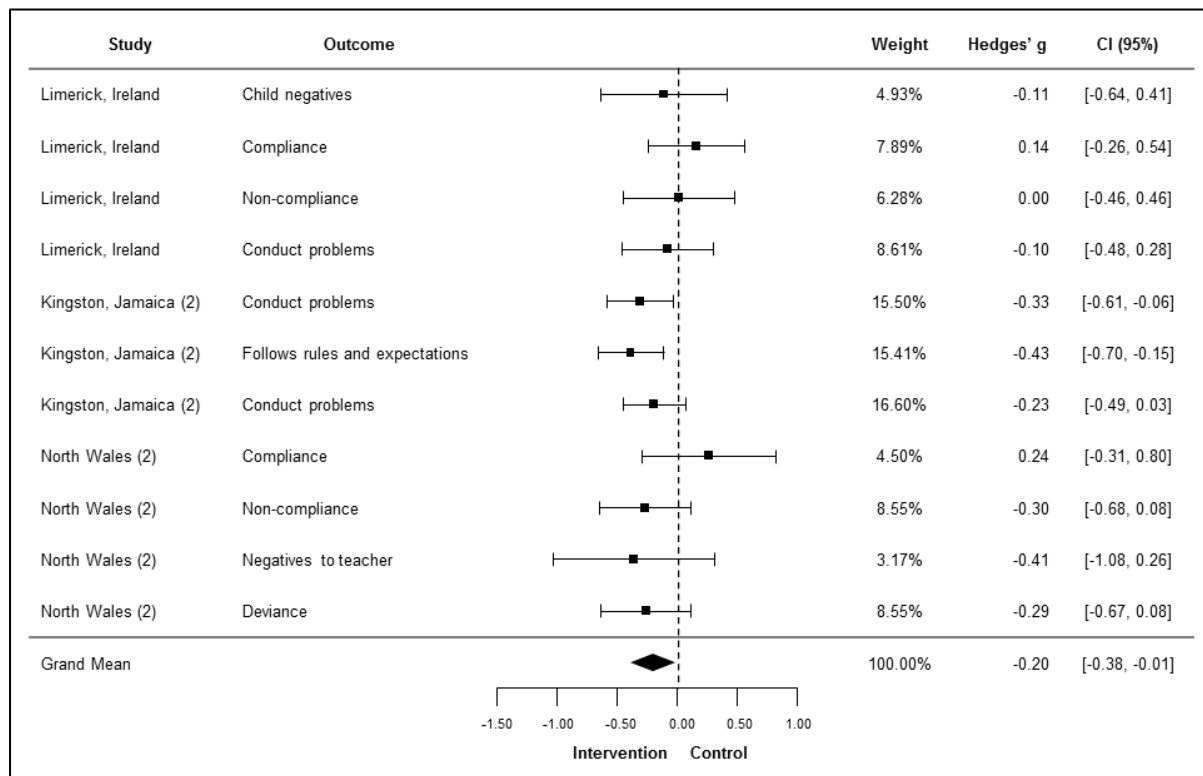


Figure 3.5 Multilevel meta-analysis of child conduct problems

Because the Limerick and North Wales (2) studies reported data on a cross-section of children with baseline teacher-reported rates of behaviour difficulties and the Kingston (2) study only reported data on those children with high levels of baseline behaviour difficulties, a sensitivity analysis examined the overall effect without the Kingston (2) data. This analysis of eight effect sizes from the two studies included data from 324 children (163 intervention, 161 control) using the T-POT and SDQ measures. The resulting effect size was not significant ($g = -0.11$, $[-0.32, 0.09]$, see Figure 3.6).

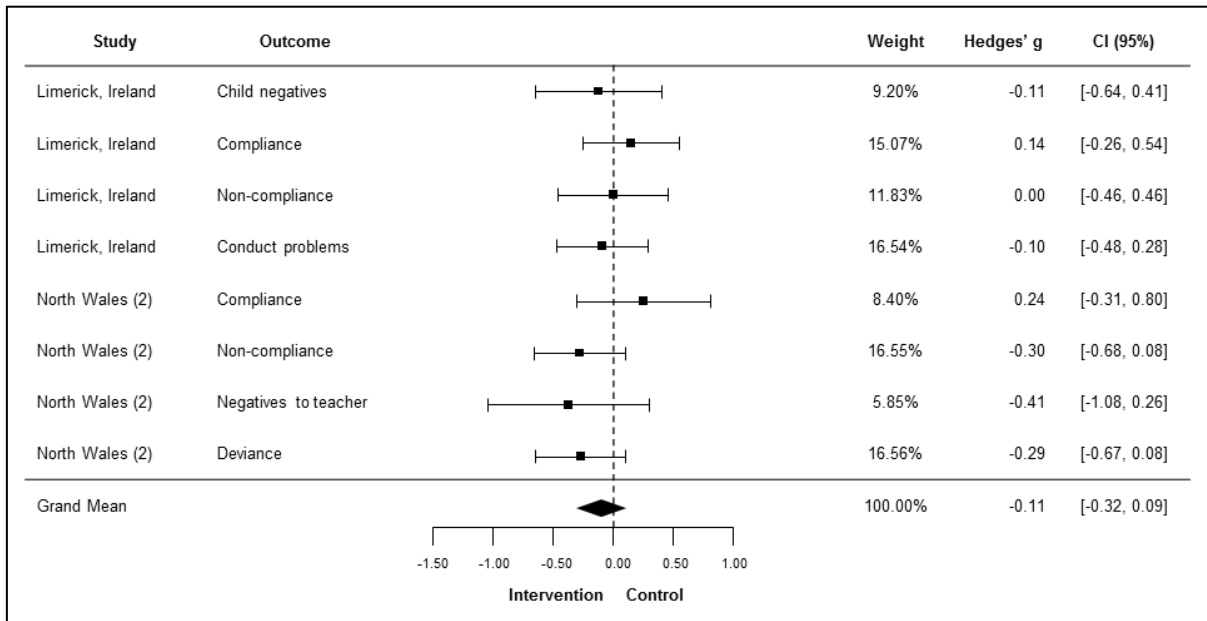


Figure 3.6 Sensitivity analysis of child conduct problems

The analysis of the high-risk subgroup contained data from three studies (Limerick, Kingston (2), North Wales (2)), involving 315 children (159 intervention, 156 control) based on the SDQ, DIPICS/MOOSEs, and T-POT measures. There was a small, statistically significant effect of the IY TCM intervention on reducing high-risk child conduct problems ($g = -0.33$, $[-0.47, -0.18]$, see Figure 3.7).

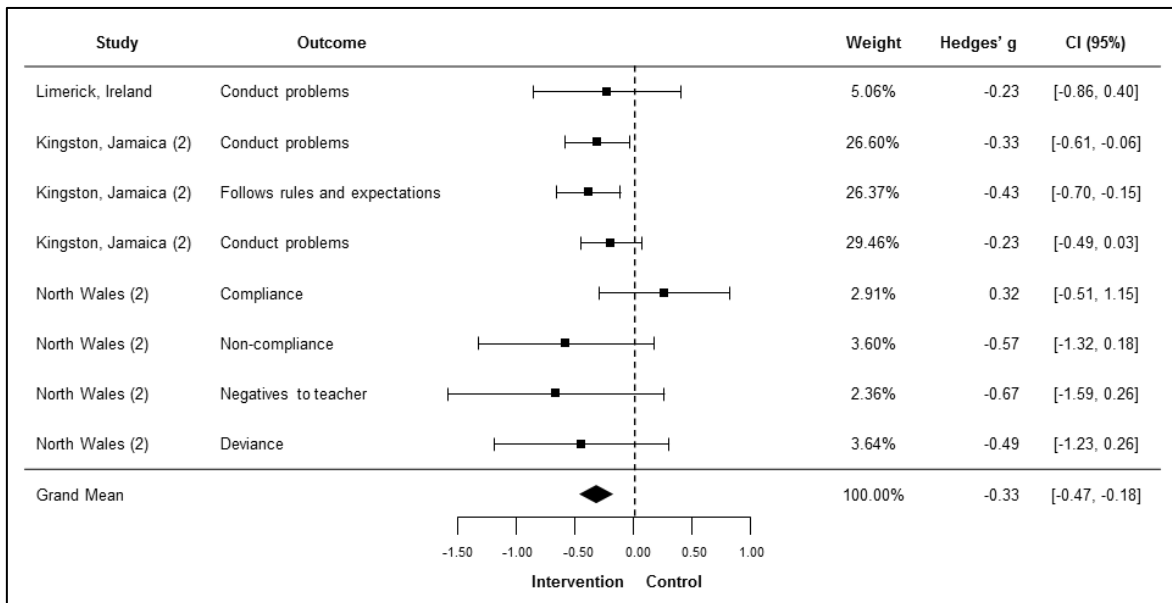


Figure 3.7 Multilevel meta-analysis of high-risk child conduct problems

Secondary outcomes.

Teacher collaboration with parents. None of the included studies reported data on teacher collaboration with parents.

Child prosocial behaviours. Child prosocial behaviours were evaluated by independent observers, teachers, and parents. Four studies reported data on teachers' strategies (Limerick, Kingston (2), North Carolina, North Wales (2)) using the T-POT, the SDQ, or the Revised Teacher Social Competence scale (R-TSC) (Conduct Problems Prevention Research Group (CPPRG), 1995). The four studies contributed six effect sizes on child prosocial behaviours for a sample size of 1,707 children (874 intervention, 833 control). There was no statistically significant effect of the IY TCM intervention on increasing child prosocial behaviours ($g = 0.18$, $[-0.04, 0.40]$, see Figure 3.8).

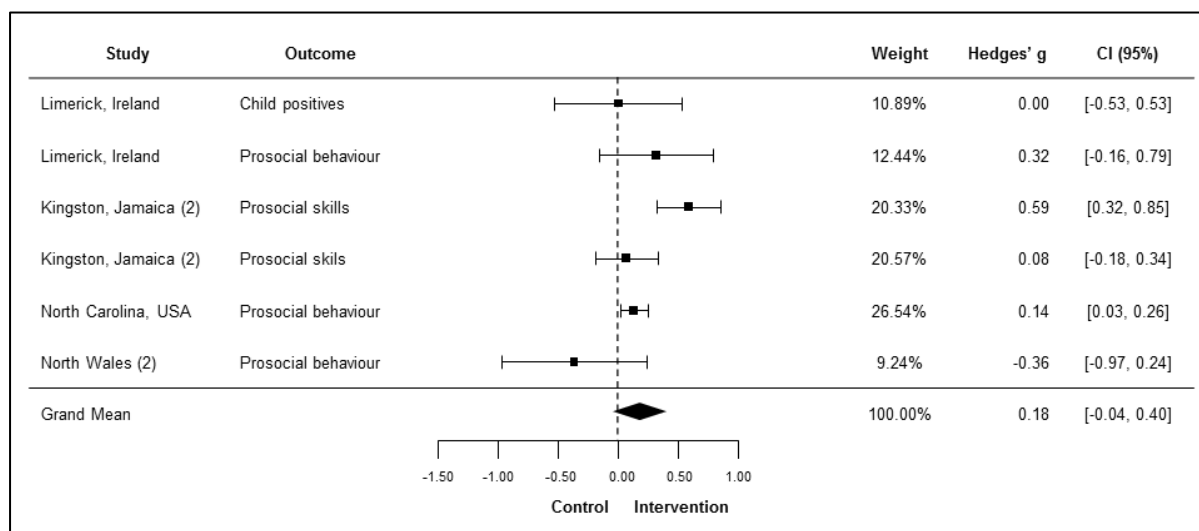


Figure 3.8 Multilevel meta-analysis of child prosocial behaviours

Again, a sensitivity analysis examined the overall effect without the Kingston (2) data. This analysis of four effect sizes from three studies included data from 1,482 children (761 intervention, 721 control) using the T-POT, SDQ, and R-TSC measures.

There was a small, statistically significant effect of the IY TCM intervention on increasing child prosocial behaviours ($g = 0.13$, [0.02, 0.24], see Figure 3.9).

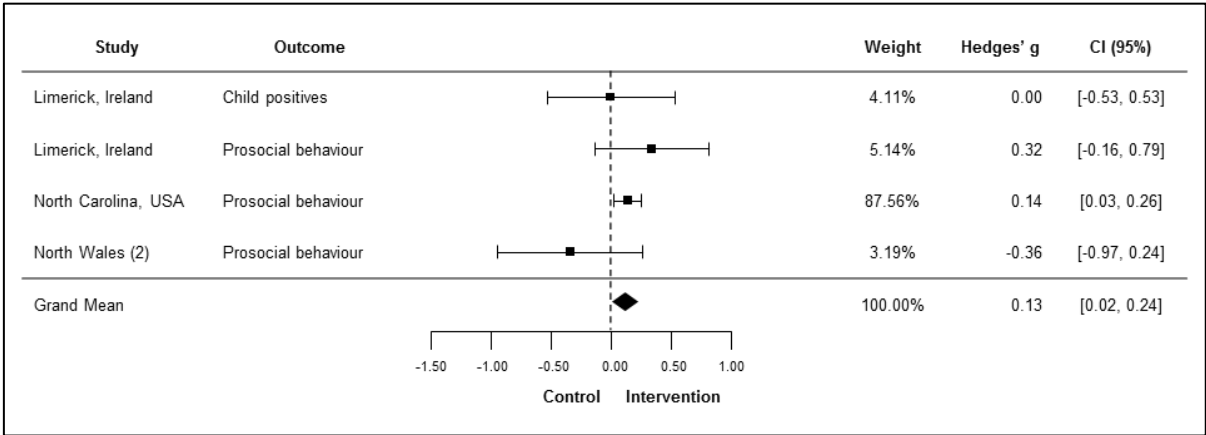


Figure 3.9 Sensitivity analysis of child prosocial behaviours

The analysis of the high-risk subgroup contained data from three studies (Limerick, Kingston (2), North Wales (2)), involving 315 children (159 intervention, 156 control) based on the SDQ, and T-POT measures. There was no statistically significant effect of the IY TCM intervention on increasing high-risk child prosocial behaviours ($g = 0.31$, [-0.03, 0.65], see Figure 3.10).

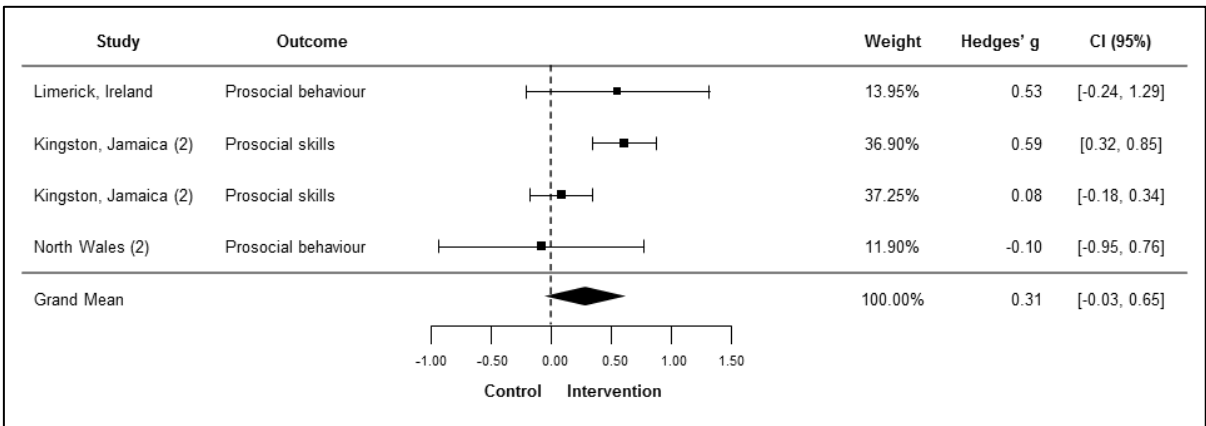


Figure 3.10 Multilevel meta-analysis of high-risk child prosocial behaviours

Child academic readiness. Only one study (North Carolina) reported on child academic readiness, so a multilevel meta-analysis was not possible for this outcome. Based on communication with the author, these data were only preliminary results of the IY TCM intervention on child academic readiness and have since been re-analysed for the final project report. As a result, until the final report has been published there is no data on child academic readiness.

Post hoc analysis: Child behaviour difficulties. As stated in the method section, not all effect sizes for child conduct problems isolated conduct problems from other constructs, such as hyperactivity and peer problems. Therefore, this post hoc analysis examines the effect of the IY TCM intervention on child behaviour difficulties, operationalised as conduct problems, hyperactivity, and peer problems. See Table 3.5 for a summary of these post hoc multilevel meta-analysis results.

Table 3.5 Multilevel meta-analysis results for post hoc outcomes

Outcome	N(k)	β_{00}	SE	95% CI	I^2	Tau^2	Chi^2 (<i>p</i> -value)
Post hoc analysis							
Child behaviour difficulties	21(4)	-0.20*	0.08	(-0.35, -0.04)	L1: 0.00 L2: 0.00	L1: 0.00 L2: 0.02	26.89 (0.14)
Sensitivity	13(3)	-0.12*	0.06	(-0.23, -0.01)	L1: 0.00 L2: 0.00	L1: 0.00 L2: 0.00	9.13 (0.69)
High risk subgroup	17(3)	-0.37***	0.05	(-0.46, -0.28)	L1: 0.00 L2: 0.00	L1: 0.00 L2: 0.00	13.63 (0.63)

N=number of effect sizes (Level 1); *k*=number of studies (Level 2); β_{00} = Hedges' *g* ('Grand Mean');
p*<0.05, *p*<0.01, ****p*<0.001

Independent observers and teachers evaluated child behaviour difficulties at post-intervention. Four studies reported data on behaviour difficulties (Limerick, Kingston (2),

North Carolina, North Wales (2)) using the T-POT, the SDQ, the DPICS/MOOSSES, Sutter-Eyberg Student Behaviour Inventory (SESBI) (Eyberg & Pincus, 1999), Conners' Inattention (Conners, 2001), Conners' Global Index (Kollins, Epstein, & Conners, 2004), R-TSC, or Pre-kindergarten Behaviour Scale (PKBS) (Merrell, 1996). The four studies contributed 21 effect sizes on child behaviour difficulties for a sample size of 1,707 children (874 intervention, 833 control). There was a small, statistically significant effect of the IY TCM intervention on reducing child behaviour difficulties ($g = -0.20$, $[-0.35, -0.04]$, see Figure 3.11).

A sensitivity analysis examined the overall effect without the Kingston (2) data. This analysis of 13 effect sizes from three studies included data from 1,482 children (761 intervention, 721 control) using the T-POT, SDQ, Conners' Scale, and R-TSC measures. There was a small, statistically significant effect of the IY TCM intervention on reducing child behaviour difficulties ($g = -0.12$, $[-0.23, -0.01]$, see Figure 3.12).

The analysis of the high-risk subgroup contained data from three studies (Limerick, Kingston (2), North Wales (2)), involving 315 children (159 intervention, 156 control) based on the SDQ, DPICS/MOOSSES, SESBI, T-POT, Conners' Scale, and PKBS measures. There was a small, statistically significant effect of the IY TCM intervention on reducing high-risk child behaviour difficulties ($g = -0.37$, $[-0.46, -0.28]$, see Figure 3.13).

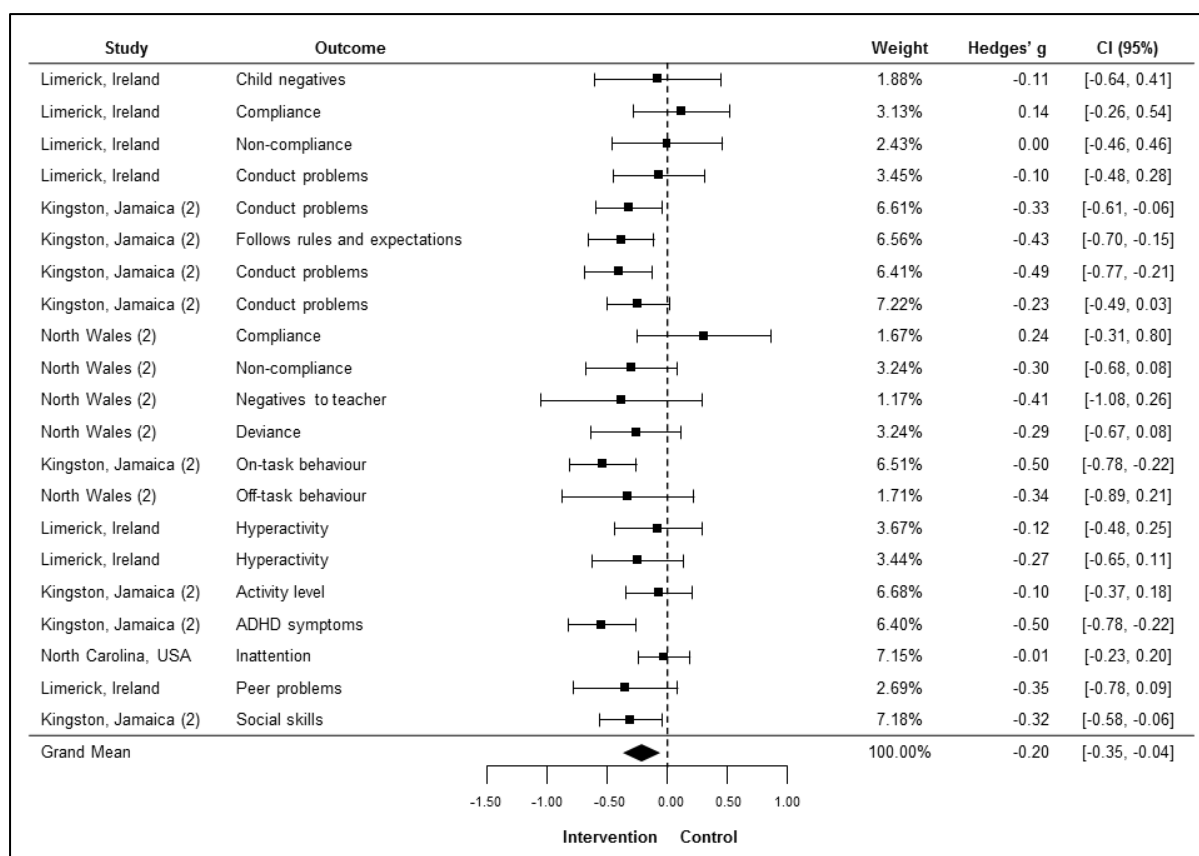


Figure 3.11 Multilevel meta-analysis of child behaviour difficulties

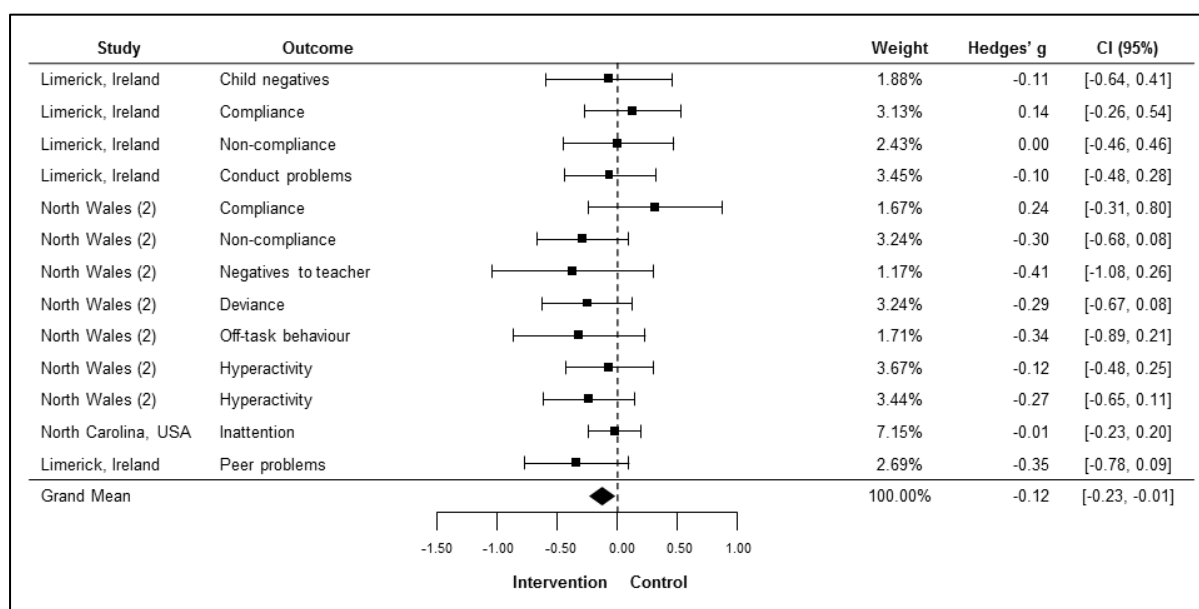


Figure 3.12 Sensitivity analysis of child behaviour difficulties

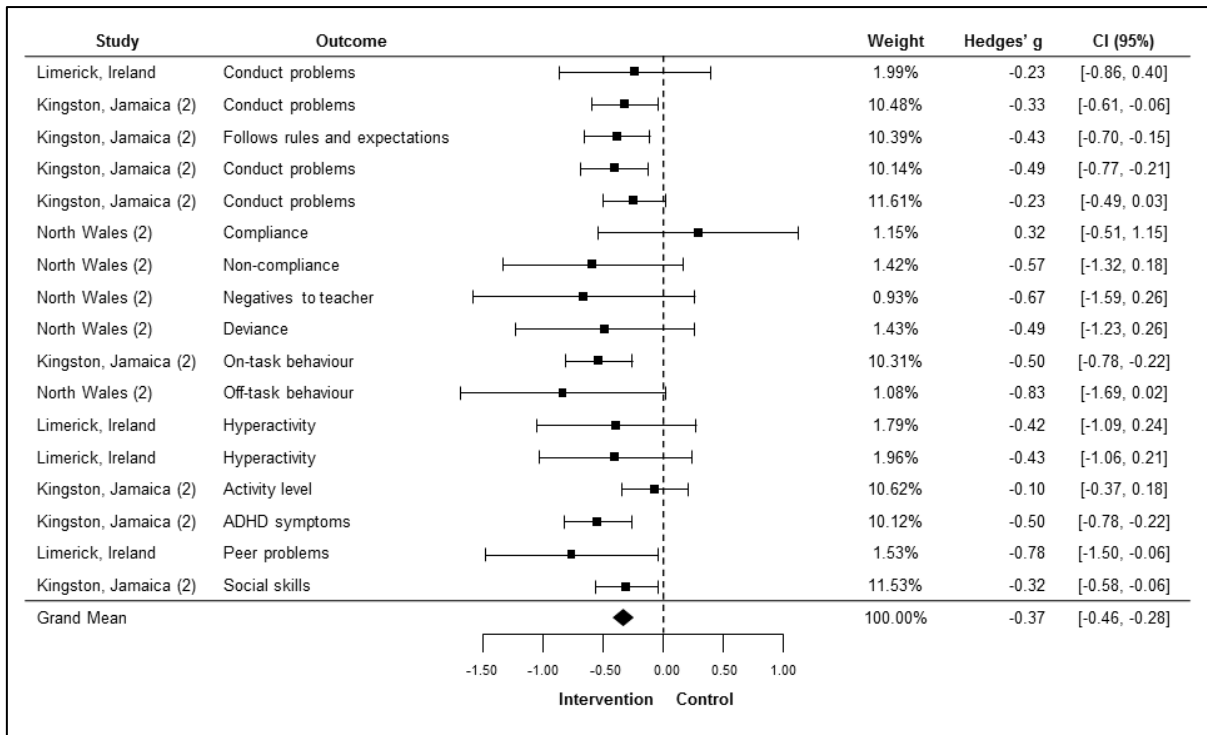


Figure 3.13 Multilevel meta-analysis of high-risk child behaviour difficulties

Discussion

Summary of main results. Overall, the results from this systematic review and multilevel meta-analysis suggest that the IY TCM programme has resulted in change at the teacher and child levels. There were small effects of the IY TCM intervention on reducing teacher negative classroom management strategies, child conduct problems, and high-risk child conduct problems and moderate effect of the IY TCM intervention on increasing teacher positive classroom management strategies. Where results were not statistically significant, patterns of effect sizes still trended towards the desired direction. Given how few included studies were involved in this multilevel meta-analysis, it is unsurprising that confidence intervals crossed zero and some effects were not significant. However, it could also be argued that given the few number of studies, clear trends are emerging to suggest that the IY TCM programme does indeed affect teacher and child outcomes in the desired direction. It is anticipated that the addition of more studies will serve to further refine

these estimates, and with continued low heterogeneity between studies, add power to the computations. In areas such as classroom environment and child academic readiness multilevel meta-analysis was not possible and individual study findings were either inconclusive or too preliminary to be able to draw conclusions. Child behaviour difficulties as a construct that incorporated a broader definition of problematic behaviour to include inattention/hyperactivity and peer problems followed the same patterns of effect sizes as all *a priori* outcomes, in that there was a small effect for all children, with the effect for high-risk children being three times greater.

Quality of evidence. None of the studies was at high risk of bias, and overall the quality of evidence included in this systematic review and multilevel meta-analysis of effectiveness trials was high. In particular, the Limerick, Kingston (2), and North Wales (2) studies were well-reported and indicated that the designs were robust and minimised potential for bias. The Rhode Island and North Carolina studies were reported in conference documents and therefore it follows that there was less information upon which to make decisions of risk of bias. Even still, there were no explicit concerns raised to question the applicability of the evidence.

Potential biases in the review process. This review is not without its limitations. Although the search terms were piloted and electronic databases were searched for ‘all fields’ rather than limited to keywords or titles/abstracts, it is possible that unpublished studies or reports not indexed in the main electronic databases could exist.

Additionally, ‘treatment-as-usual’ is not a standardized control across included studies, unlike the IY TCM intervention. In conducting a multilevel meta-analysis that compares effect sizes calculated from standardised means, the analysis is in fact answering the question, ‘Is the IY TCM Programme more effective than treatment-as-usual?’

Without each included study explicating the components included in their respective

control groups, it is not possible to establish how theoretically comparable the comparison conditions are. This limitation is relevant to all meta-analyses, but to the extent that authors do not over-generalise the implications that can be drawn from their studies this approach still provides us with the best information on programme effectiveness.

Agreements and disagreements with other reviews. The Oliver et al. (2011) review concluded that teacher classroom management programmes do have a significant overall effect on decreasing problem behaviour in the classroom. Narrative synthesis of the 14 included studies in the Whar et al. (2012) review revealed inconsistent findings of teacher management on children's social and prosocial skills. There were trends towards decreased problem behaviours due to intervention, and teacher positive management strategies were found to be moderately statistically significant in favour of intervention. Overall, this current review and multilevel meta-analysis has produced results that further support the effectiveness of teacher-training as an intervention to benefit children's behaviour.

Conclusion

Implications for practice and policy. The results from this systematic review and multilevel meta-analysis of effectiveness trials of the IY TCM programme demonstrate the benefits possible in improving teachers' strategies and children's problematic behaviours. It is notable that the results of this analysis represent only those studies that implemented the IY TCM programme in isolation from all other interventions. By incorporating data from all reported measures (e.g., teacher reports, independent observations), a more comprehensive estimation of the outcomes was produced, minimising the limitations inherent in any one type of measure alone. Therefore, the study design and analysis strategy allow for stronger conclusions to be drawn about how schools can effectively address problematic child behaviours in the classroom through teacher-training alone.

Implications for research. There is still a dearth of quantitative evidence on the IY TCM programme, and what data that does exist has been collected within a handful of countries and with limited long-term follow-up. Therefore, future research is needed to address moderate- and long-term results of teacher-training programmes. Including studies from a broader variety of countries (specifically low- and middle-income countries and non-Western countries) would provide valuable information about how this intervention translates across contexts. While this review isolates the IY TCM programme from other teacher classroom management programmes, there is room to expand the multilevel meta-analysis to include other interventions. Expanding this review to include data from studies evaluating other teacher-training programmes that are designed around the same theoretical tenets and operate within similar logic models could provide a broader understanding of the teacher and child benefits expected from teacher classroom management strategies.

Chapter 4 : Experiences of the IY TCM programme from key stakeholders – Results of a qualitative systematic review and meta-synthesis

Results

To avoid repetition, details of the results of the search can be found in the previous chapter, before detailing the results of the quantitative strand of this systematic review.

Description of studies

Included studies. As reported in Chapter 3, nine studies in 14 records met the inclusion criteria for this mixed methods review (see Table 4.1). This section highlights the characteristics of all included studies in the qualitative strand (see Table 4.2) to provide context for the subsequent findings from the qualitative meta-synthesis. Five studies reported in seven records were included in this qualitative strand. Two of these studies took place in Ireland and one each in England, Jamaica, and Wales.

Table 4.1 Included studies

Included studies for Qualitative Strand	
Study ID	Records
Devon, England	Marlow et al., 2015
Greater Dublin Area, Ireland	Davenport & Tansey, 2009
Limerick, Ireland	McGilloway et al., 2011
	McGilloway et al., 2012
	Hyland, 2014
Kingston, Jamaica (1)	Baker-Henningham & Walker, 2009
North Wales (1)	Hutchings et al., 2007

Devon, England Study. This mixed methods study combined qualitative focus groups, semi-structured interviews, and standardised quantitative questionnaires from a pre-post study to explore the feasibility of implementing the IY TCM programme in schools as a public health intervention. Only the qualitative findings met inclusion criteria for this review and were included. The findings from this study are reported in one peer-

reviewed journal article (Marlow et al., 2015). This study aimed to examine the experiences of participating teachers and school head teachers as well as determine the feasibility of recruiting schools and implementing a cluster randomised controlled trial of the IY TCM programme across schools.

Forty class teachers had the opportunity to participate in the IY TCM training over the course of two years, and 37 completed the course. Reasons for the three teachers who did not complete the course included: leaving the teaching profession, feeling the course was too basic, and school staffing issues. About a month after the final training session occurred, teachers were invited to participate in focus groups (in same groups as their training sessions).

Thirty-one of the 37 class teachers attended the focus groups, which lasted no more than one-and-a-half hours and were led by a researcher, with an observer also in attendance. The semi-structured session ran through questions pertaining to the IY TCM ‘course content, style and mode of delivery, theory, practical skills, materials, group learning, arrangements for attending and feasibility, dissemination within their school and other programmes used within schools’ (Marlow et al., 2015, p. 235). The focus groups also covered the experience of participating in a research study and suggestions for improvements for the main research trial.

Sixteen head teachers (one from each participating school) took part in semi-structured telephone interviews after their staff members had completed the course. These interviews lasted no more than 45 minutes and were structured to cover general school information, feedback from their staff members who participated in the IY TCM training, feasibility, and discussion of any similar (but non-IY TCM) programmes currently operating in their schools.

All focus groups and telephone interviews were audio-recorded and transcribed in anonymous form. Researchers applied the framework approach to thematically analyse the data deductively from the topic guide structure while still allowing for emerging themes to develop from the data.

Greater Dublin Area, Ireland Study. This mixed methods study combined a qualitative survey evaluating the IY TCM programme with quantitative questionnaires on teacher and child outcomes following a pre-post study (no control). One report (Davenport & Tansey, 2009) discusses these findings, and only the qualitative data meets inclusion criteria for this review. The 15 participating teachers came from rural and suburban schools, and were asked to complete a qualitative survey. Results from these questionnaires were analysed by survey headings and reported with frequency counts using ‘quasi-statistics’ (e.g., ‘all’, ‘some’). Topics included IY TCM approach to classroom management, practical components of the course, materials, and delivery.

Limerick, Ireland Study. This mixed methods study combined the experimental design of a cluster randomised controlled trial across 11 schools with 22 teachers and 217 index children and a two-part qualitative process evaluation with class teachers. This study’s findings are detailed across four papers, including two reports (McGilloway et al., 2010; McGilloway et al., 2012), one PhD dissertation (Hyland, 2014), and one peer-reviewed journal article (Hickey et al., 2015). All records except for the 2015 journal article were included in this qualitative strand.

For the qualitative process evaluation, researchers conducted one-to-one semi-structured interviews with a subgroup of participating teachers immediately post-intervention and then at follow-up 12-months post-baseline. Five intervention teachers participated in the first post-intervention interviews, and six control teachers participated the following year immediately after they received the IY TCM training due to the waitlist

design. These 11 teachers discussed their experiences participating in the IY TCM training programme and their classroom management strategies after course completion. Additional interviews were conducted six months after post-intervention (i.e., 12-months post-baseline) with six teachers, who discussed their perceptions of the learned strategies after long-term use and related topics, including whether the techniques were similarly applicable with their new classroom of children (it was the following academic year). With both the post-intervention and follow-up interview data, audio-recordings were transcribed for thematic analysis using a framework analysis approach.

Kingston, Jamaica (1) Study. This qualitative study is reported in one peer-reviewed journal article (Baker-Henningham & Walker, 2009). The aims of this study were fourfold: to examine teacher perceptions of the IY TCM programme, to explore barriers and facilitators to implementing IY TCM strategies in the classroom, to understand teacher views on the acceptability of the IY TCM programme, and to hear teacher beliefs on whether the IY TCM programme was effective. The 15 participating teachers came from three pre-schools (urban). In this context, class teachers generally are qualified at the paraprofessional (i.e., teaching assistant) level. As such, this sample included 12 teachers who had finished a vocational early childhood course, with only one of the participants classified as a ‘trained teacher’. The intervention implemented in this study was the IY TCM programme with some aspects of the IY Dina Dinosaur Classroom Curriculum, but for the purposes of this review, only the qualitative data relating to the IY TCM programme is included. Following completion of the intervention programme, independent researchers (i.e., researchers who had not previously been involved in the study) conducted in-depth individual interviews with participating teachers. Each interview lasted between 45 minutes and one hour in a quiet place at the teachers’ schools. Interviews followed a semi-structured topic guide, were audio-recorded, and transcribed

for analysis. Researchers applied the framework approach for analysis to code themes and categories before charting data underneath each theme to find patterns. Researchers did note the number of teachers who reported each theme as an indication of salience within the data.

North Wales (1) Study. This mixed methods pre-post study is reported in one peer-reviewed journal article (Hutchings et al., 2007), although only the qualitative data is included in this review. Twenty-one teachers (out of 23 IY TCM participants) took part in semi-structured interviews to discuss topics including IY TCM content, implementing IY TCM strategies in the classroom, and recommendations for the research team. Researchers applied thematic content analysis to the transcripts of these interviews. Two independent coders read through the transcripts and applied seven categories to the data, and the authors report that inter-rater reliability was high (Kappa's coefficient = 0.82).

Excluded studies. Twelve records were examined and excluded from the qualitative strand of this review for failing to meet all of the a priori inclusion criteria set forth (see Table 4.3). Most records were excluded because they were reviews of the literature and not primary qualitative studies.

On-going studies. One study was classified as on-going for the qualitative strand of this review, even though it was also included in this review. The Devon, England Study has published initial qualitative findings, and yet because it is not fully completed, it is expected that more qualitative findings (and quantitative results) will be published. As previously mentioned, this study's main protocol (Ford et al., 2012) and process evaluation protocol (Hansford et al., 2015) describe a large, cluster-randomised controlled trial occurring over three cohorts of participating teachers/schools and a two-year follow-up period.

Table 4.2 Characteristics of included studies

Study ID	Participants	Method	Topics	Analysis
Devon, England	31 teachers and 16 headteachers from state-funded primary schools	Teachers Focus groups comprising intervention training group members, one month post-intervention Headteachers Semi-structured telephone interviews, post-intervention	Teacher focus groups Course content, delivery, theory, group learning, skills and materials, feasibility, dissemination, and research process Headteacher interviews General school information, feasibility, similar programmes, and teacher feedback about training	Thematic analysis and Framework approach (coding frameworks based on topic guides but also developed iteratively during analysis)
Greater Dublin Area, Ireland	15 teachers	Written survey, post-intervention	Teacher surveys Confidence managing child behaviours, recommendation of IY TCM workshops to colleagues, endorsement of IY TCM content, perceived improvement in child behaviours, perceptions of course materials, timing of course, perceived usefulness of course strategies	Frequency counts based on topic guide headings
Limerick, Ireland	11 teachers (post-intervention) from 6 schools 6 teachers (1 year post-intervention follow-up)	One-to-one, in-person semi-structured interviews	Post-intervention interviews Current classroom management issues, strategies used to deal with classroom difficulties, participation experiences, post-training classroom management 1-year follow-up How do teachers' experiences change over time, perceived longterm effectiveness, best aspects of programme and any barriers	Thematic analysis and Framework approach
Kingston, Jamaica (1)	15 teachers from 3 community pre-schools	One-to-one, in-person in-depth interviews	Post-intervention interviews Opinion of training programme and perceptions of benefits, how teachers used the strategies and most/least useful strategies, views on how intervention brought about changes, suggestions for future implementations	Framework approach
North Wales (1)	21 teachers (IY TCM courses had been delivered in both Welsh and English)	One-to-one interviews	Post-intervention interviews Background information, IY TCM course, classroom application of skills learned, practical implications, comments and suggestions for future	Thematic content analysis

Table 4.3 Excluded studies

Record	Reason for exclusion
Considered only for qualitative strand	
Baker-Henningham, 2011	Pre-intervention interviews
Maras, Splett, Reinke, Stormont, & Herman, 2014	Not specific to IY TCM
Considered both for qualitative and quantitative strands	
Colletta, Latimore, Reinke, & Herman, 2013	Informational study brochure
Hutchings, 2012	Programme report
Pidano & Allen, 2015	Literature review
Webster-Stratton, 2001	Literature review
Webster-Stratton & Bywater, 2014	Literature review
Webster-Stratton & Herman, 2010	Programme development and fidelity
Webster-Stratton & Reid, 2007	Literature review
Webster-Stratton & Reid, 2010	Literature review
Webster-Stratton, Gaspar, et al., 2012	Literature review
Webster-Stratton, Reinke, Herman, et al., 2011	Discussion on programme fidelity

Quality appraisal for included studies

Excerpts from the studies' reports are provided in Appendix E to support the quality assessment ratings applied to each study and described below (also see Figure 4.1). The interrater percent agreement across all items assessing quality in the included studies was 70 per cent. Where differences existed, the two raters discussed and reached agreement on a final assessment.

Rigour of sampling. Rigour of sampling was determined based on three criteria: sampling strategy, attempts to obtain a diverse sample, and reported characteristics of the sample. The manner by which a sample was obtained has important implications for ensuring that the findings are supported by the intended group of interest (Tong, Flemming, McInnes, Oliver, & Craig, 2012; Tong, Sainsbury, & Craig, 2007). Qualitative studies do not aim for statistical generalizability; instead, they are interested in theoretical generalizability, which is why appropriate sampling techniques (e.g., purposive) and

explanations of rigour in conducting the sampling are important. Overall, the studies in this review had high levels of rigour in sampling, with one study (Limerick) in particular deemed to have made a thorough attempt to increase the rigour in the sampling strategy. Three studies (Devon, Kingston (1), North Wales (1)) were rated as having taken several steps to increase the rigour in sampling. Only the Dublin study was rated as medium quality due to lack of clarity in how many steps were taken to increase the rigour in sampling.

	Rigour of sampling	Rigour of data collected	Rigour of analysis	Findings grounded in data	Breadth and depth	Overall reliability and trustworthiness	Overall usefulness of findings for this review
Devon, England	High quality	Medium quality	Medium quality	High quality	Medium quality	Medium quality	High quality
Greater Dublin Area, Ireland	Medium quality	Low quality	Low quality	Low quality	Low quality	Low quality	Low quality
Limerick, Ireland	High quality	High quality	High quality	High quality	High quality	High quality	High quality
Kingston, Jamaica (1)	High quality	Medium quality	Medium quality	High quality	Medium quality	High quality	High quality
North Wales (1)	High quality	Unclear quality	Unclear quality	Medium quality	Medium quality	Medium quality	Medium quality

Figure 4.1 Quality appraisal of included studies

Rigour of data collected. Rigour of data collection was assessed on three criteria: piloting of collection tools (e.g., topic guide), comprehensiveness and flexibility, and ensuring all participants were able and willing to contribute. One study (Limerick) was rated as having taken several steps and was rated as high quality. Two studies (Devon, Kingston (1)) were rated as of medium quality for having taken a few steps to increase the rigour in the data collection. It was unclear for the North Wales study what steps were taken, and the Dublin study was similarly unclear and low in quality on rigour in data collection.

Rigour of analysis. Four criteria guided decisions on rigour in data analysis: description of methods, diversity in perspective, balance in analysis, and ruling out of alternative explanations. Three studies were rated as having taken a few steps to increase rigour in the analysis of data (Devon, Limerick, Kingston (1)). It was unclear what steps were taken in one study (North Wales (1)) and deemed not sufficiently rigorous in another (Dublin).

Findings grounded in data. The extent to which a study grounded its findings in the data was based on four criteria: whether enough data were presented, if the data presented fit the interpretation, if it illuminated the findings, and whether quotations were numbered. Use of quotations in reporting qualitative research grounds the findings in the data for readers and makes it easier to understand the results (Côté & Turgeon, 2005). Providing identification numbers alongside quotations enhances transparency of the reporting and boosts the trustworthiness of the findings that the narrative is not driven by one or few cases as opposed to emerging more broadly across participants (Tong et al., 2012; Tong et al., 2007). Three studies (Devon, Limerick, Kingston (1)) were deemed to have well-grounded findings, one study (North Wales (1)) had fairly well-grounded findings, and one study (Dublin) had limited grounding of its findings.

Breadth and depth. Determining the extent to which studies had breadth and/or depth was established through four criteria: the range of issues covered, how fully explored the participants' perspectives were, the richness and complexity presented, and the inclusion of theoretical/conceptual development. One study (Limerick) was rated as having good/fair breadth and depth. Three studies (Devon, Kingston (1), Wales (1)) were rated as having good breadth but very little depth, and one study (Dublin) was deemed to have limited breadth or depth.

Overall quality ratings. As a result of the studies' ratings on the above five areas, one study stood out as having the highest overall reliability, trustworthiness of findings, and usefulness for this review: Limerick. The Devon and Kingston (1) studies also were rated moderately high across the criteria and were determined to have good reliability, trustworthiness, and usefulness for this review. The North Wales (1) study was rated of moderate quality and usefulness for this review, and only the Dublin study was rated low overall. As such, the Limerick study served as the index paper for this qualitative analysis and it spiralled out through the other higher quality papers so that the Dublin study was the last one incorporated into the analysis. Two excerpts with examples of the coding process are available in Tables 4.4 and 4.5 to illustrate that initial open codes were iteratively grouped first into sub-codes and then into codes of related constructs before being nested within the overarching themes expressed through the analysis.

Table 4.4 Example of coding tree for teacher outcomes theme

Themes	Codes	Sub-codes	Initial open codes
Teacher outcomes	Knowledge	Children's development, abilities, and needs	Theory of child development, differences in children, understand children's needs and abilities, expectations more appropriate for child behaviour, need to give all children attention and encouragement
		Classroom management theory and strategies	Theory behind management strategies, crystallise theory into practice, see how effective strategies can be like proximal praise, spending individual time with students, incorporating interactive teaching practices, developmentally appropriate strategies, practical and concrete strategies
	Locus of control	Better equipped to deal with day-to-day challenges	New ability to deal effectively, feel more confident, 'I can do this', reaffirming, licence to change practice, able to 'let go' of issues, increased self-efficacy, 'greater sense of freedom in the classroom'
		'Resource bank' of strategies for current and future use	Ignoring seen as active strategy, more strategies to choose from, making more effort to communicate with parents, find it rewarding to pass on information
	Emotional well-being	Calmer, less stressed while in the classroom	Reduced stress, calmer, seeing results, don't have to stop every minute, awareness that others are in the same situation, not isolated, developed new skills to remain calm, experience 'appropriate' emotional responses
		More pleasure and enjoyment while with children in classroom	Feel good, see children smile, have fun playing with children
	Teaching philosophy and practice	Re-enthused about teaching in general	Changed way of thinking, fundamental change, greater sense of freedom, home-school communication and collaboration, more interactive teaching methods, energising, optimism towards the future, preparedness for incoming classes
		Reliance on positive strategies	Fewer negative strategies, spending individual time with children, using developmentally appropriate strategies, proximal praise and encouragement, affirmation of positive techniques

Table 4.5 Example of coding tree for critiques theme

Themes	Codes	Sub-codes	Initial open codes
Critiques	Motivation for taking course	Compulsory versus choice	Volunteered to receive training, asked by head teacher, facilitators think voluntary participation facilitates learning, want whole-school participation
	IY TCM process and content	Scripted versus flexible	Delivery felt scripted, acknowledge variety of starting skills, teachers desire flexibility, could have covered two topics per day, facilitators wary of cutting up programme
		Role play and video vignettes	discomfort with role play, challenging aspects encouraged learning, vignettes promote group discussion, video vignettes not reflective of non-American classrooms
		Misunderstandings and disagreements	Challenge to ignore aggressive behaviour, see value in corporal punishment
	Child outcomes	Age and other factors	Examples more relevant for younger children
		Not meant for particularly complex children	Complex children awaiting psychological assessment won't benefit, harder to use strategies when including children with SEN
	Wider support and scaling-up	Investment from school leadership	Head teachers'/school principals' support influences teacher engagement, need school leadership support
		Training rolled out to all teachers and school staff	Short-lived effects without whole school implementation, inconsistency among other school staff problematic, need to train TAs, playground staff, canteen staff, national roll-out for pupil and teacher benefits, share strategies with other teachers, would recommend colleagues to take course, possibly incorporate into BEd system, particularly needed at start of teaching career, share strategies with parents

Experiences of key stakeholders

Analysing the findings from each included study in this qualitative strand produced an overarching understanding of how key stakeholders (predominantly classroom teachers) experience the IY TCM programme (see Figure 4.2). What emerged was a cyclical process of engaging in the training and observing effective outcomes. First, the cycle of learning, or the teachers' iterative process of attending the group training sessions and then individually practicing learned strategies in the classroom before the next month's group training session, will be detailed. Over the period of the entire course, the teachers cycled between primary learning opportunities (when they learned firsthand from their own experiences) and secondary learning opportunities (when they learned through others' experiences). Ultimately, this led to their experiences of the effects of the IY TCM training, the second part of this meta-synthesis. More specifically, outcomes described by the stakeholders are explained as they pertain to the teachers, the classrooms, and the children. This structuring reflects the hierarchical relationship existing in the educational

environment, in which the children are nested within the teacher/classroom level, and the IY TCM intervention experience, with the teachers receiving the direct training to indirectly affect the classroom and children.

Cycle of learning. The cycle of learning emerged not only from the design of the IY TCM course but also in the process experienced by the teachers as they progressed throughout the academic year. The intersection of the individual and the group is evident from the very beginning, and the IY TCM facilitators (group leaders) are in the crucial position of ensuring the smooth integration of each teacher into the training group.

‘[...] the professionalism and competence of the TCM group facilitators was an essential component of training here. They were also appreciated by the participants; indeed, seven teachers specifically noted the role that facilitators played in their learning experience, and commented on their approachability, commitment, organizational skills, and their expertise.’ (Hyland, 2014, p. 167)

‘All teachers reported having developed strong relationships with the group leaders.’ (McGilloway et al., 2010, p. 17)

‘Teachers generally liked the collaborative, responsive approach demonstrated by the group leaders and the fact that they used their own experiences as teachers as well as recognizing and drawing on the skills of the participants.’ (Marlow et al., 2015, p. 240)

Even when teachers felt that the delivery did not match their needs all of the time, they were still positive about the role played by the facilitators.

'Teachers were positive about leaders modelling different techniques [...] Some teachers expressed a view that the level of delivery was sometimes pitched too low and that they felt there was a "script" that the leaders had to follow which could have been more flexible. However, some thought that the leaders got the level of delivery just right and others acknowledge the need for balance to meet all needs.' (Marlow et al., 2015, p. 240)

Certainly, adjusting delivery of the course to meet all needs is challenging because all teachers possess their own set of knowledge and skills about children's learning and classroom behaviour management when they begin the training course. The benefit of bringing together a group of teachers, who typically operate independently in their own classrooms, is in the shared learning and sense of camaraderie that develops over the period of training. In this way, the various levels of knowledge, skills, and teaching experience that present a challenge to the facilitators also present an advantage to participating teachers.

Throughout the monthly training sessions and individual practice between sessions, teachers learned through their own experiences and reflection of practice (primary learning) and through the shared experiences and collective reflection of their IY TCM group (secondary learning). To start, each teacher arrived with previous knowledge and skills, and initial group relationships needed to be established and a positive group dynamic fostered.

'All reported knowing and using the theoretical underpinnings of TCM strategies prior to the training, but regarded the programme as building upon what they knew and were trying to achieve. According to their responses, the course had provided them with an invaluable opportunity to crystallise their theoretical knowledge into conscious practical applications in the classroom.' (McGilloway et al., 2010, p. 17)

‘The way in which the groups were facilitated, the group size and the time for teachers to get to know each other over a longer period of time helped to create a mutually supportive, trusting environment. This was valued because teachers liked being able to share difficulties in a non-judgemental environment (reducing isolation) and they felt they could try out new techniques, fail and try again.’ (Marlow et al., 2015, p. 242)

With the establishment of a safe space to learn, facilitators were able to deliver the course content, and the teachers were able to practice skills before returning to their classrooms to implement the strategies in real-world scenarios.

‘The mode of programme delivery, which allowed for the implementation of learned strategies in the classroom over the course of one month, before feeding back to the group, was considered to be a very positive element of the programme; teachers valued the opportunity to try out the different strategies in the classroom over time.’ (McGilloway et al., 2010, p. 17)

‘Teachers seemed to value the time and opportunity to practice skills and reflect between monthly sessions.’ (Marlow et al., 2015, p. 239)

‘Teachers commented on the value of having time out of the classroom to reflect on their practice, test their own assumptions, think about and discuss theoretical issues related to practice with other professionals and consolidate existing knowledge’ (Marlow et al., 2015, p. 242).

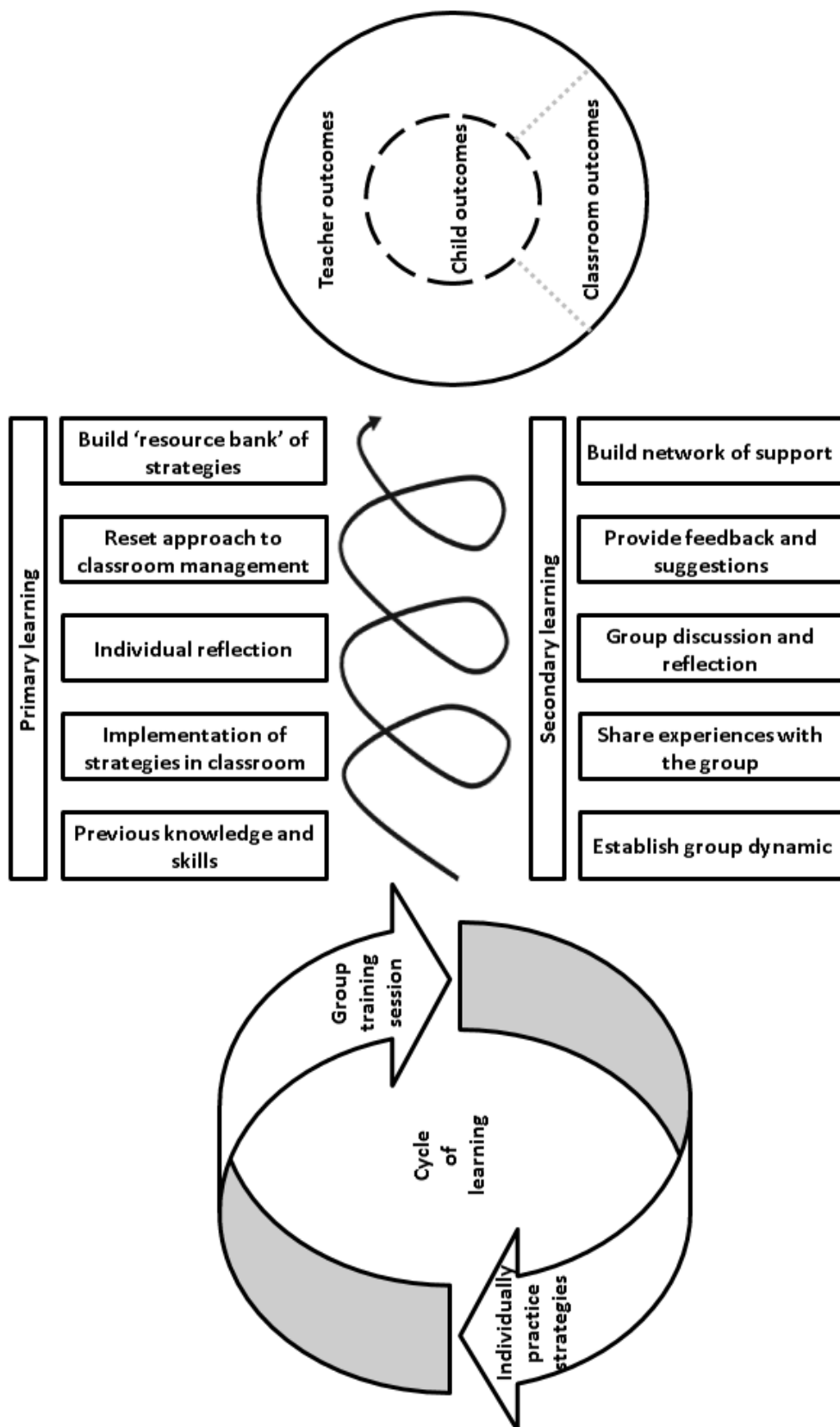


Figure 4.2 Result of the meta-synthesis: Cycle of learning

It was not just the participating teachers and facilitators who thought this training's timing allowed for professional development, but head teachers also expressed positive views of the monthly session format.

'Some heads also commented that the length of the course and the time between sessions was helpful to embed learning and observe impact within the school.' (Marlow et al., 2015, p. 240)

Even with weeks of practice between sessions, implementation of IY TCM strategies in the classroom was not always smooth. In addition to their individual reflection between sessions, teachers were able to report back to the group about their experiences trying out the classroom management skills, allowing opportunities for group reflection and advice as well as encouragement to keep trying. In the context of a group with varying levels of experienced teachers, it was noted that newly qualified teachers particularly benefited from this collective exchange.

'All teachers reported that further learning had occurred via the social medium of interaction with other teachers on the course which provided them with a readily accessible 'wealth of experience'. For example, some of the teachers on the programme had several years' experience dealing with infant classes and this was highly valued as offering potential learning opportunities, particularly by the more recently qualified teachers on the course.' (McGilloway et al., 2010, p. 17)

'[...] the feeding back of information on the successes and challenges of strategy implementation, fostered open discussion and generated solutions.' (McGilloway et al., 2010, p. 17)

'Two other themes which emerged from the focus groups and interviews in relation to the TCM course were the importance of the group learning situation, and the value of having time to reflect on practice. Teachers valued the opportunity to share experiences and the interactive nature of discussions.' (Marlow et al., 2015, p. 242)

'Sharing experiences and strategies with others was the most commonly reported benefit by 17 teachers (81%), suggesting that the collaborative philosophy of shared problem-solving had been effective and valued.' (Hutchings et al., 2007, p. 20)

It is apparent that the value of the shared learning experiences emerges not only from joint problem-solving, but also in the emotional journey of moving from feeling isolated in a perceived weakness (i.e., poor classroom behaviour management) to the realisation that other teachers also experience these emotions.

'As indicated earlier, many teachers struggle alone with classroom management difficulties and feel that behavioural problems are partly attributable to their own lack of expertise. As part of the sharing process at the heart of the TCM programme, these teachers realized that others were in a similar situation to themselves and that they were no longer alone in dealing with these issues. This was regarded as an invaluable aspect of the programme.' (McGilloway et al., 2010, p. 17)

Over time the teachers expressed that their teaching practice was realigning with the principles that they had originally learned and which were being re-taught on the IY TCM course.

‘Although teachers may have recognised the importance of positive behaviour reinforcement, this did not always translate into classroom practice [...] teachers can lose their idealism and discontinue any initial use of positive management strategies (i.e., they can experience ‘washout’) when faced with the reality of the classroom [...] within the current study ‘reverse washout’ was evident in the realignment of teaching practice with theoretical knowledge and practical skills.’ (Hyland, 2014, p. 170)

This process of ‘reverse washout’ is akin to hitting the reset button to return to the factory settings. Teachers felt they were able to go back to the basics, and re-establish positive routines for classroom management that had slipped over time. The IY TCM programme served as a refresher course, in a sense, to remind teachers of the theories underpinning child development and teaching best practice, but also allowed for hands-on practice to cement these ideas into day-to-day use. As previously mentioned, the group aspect of the training was viewed very positively, for the teachers were able to establish a support network. But, the content and materials were also important, in that they could be compiled into a resource bank of classroom management strategies for teachers to use in the future. This was crucial, because, teachers receive a new class of pupils with each academic year, and the types of challenging behaviours they might have present in their classroom will inevitably differ. Although teachers might not have needed a particular strategy with one year’s class, they could need to pull that strategy out of their resource bank in a subsequent year with a new class.

‘The course materials – course notes, books and other resources – were also regarded as extremely useful tools. After training, teachers appreciated the utility of these resources and stated that they could refer back to these whenever the need arose.’ (McGilloway et al., 2010, p. 17)

'[...] 13 teachers (62%) [reported] having an extended range of strategies available to them and 10 teachers (48%) [recognised] the value of the philosophy with its emphasis on positive management strategies.' (Hutchings et al., 2007, p. 20)

Therefore, the cycle of learning involves not only the interchange between individual teachers and the training group as a whole, but also an extended feedback loop of direct learning, practice in the group, practice in the classroom, individual reflection, group sharing and reflection, re-setting one's approach to classroom behaviour management, establishing a support network, and building a resource bank of strategies for current and future use. While this is all well and good, teachers had to experience positive effects from the training to feel that it was worthwhile. Observing children's compliance and cooperation after implementing learned strategies serves to positively reinforce these techniques for the teachers. The experiences of the outcomes from training are the second part of this overarching model.

Effects of training. In addition to their experiences learning on the IY TCM course, the teachers also provided insights into their experiences of the effects of the training. These outcomes could be classified at the teacher/classroom level and at the child level. At the teacher/classroom level, a distinction will be made for effective outcomes experienced by the teachers (teacher outcomes) and effective outcomes experienced for the whole classroom and therefore including the children as a collective (classroom outcomes). Nested within the teacher/classroom outcomes are the child outcomes, which are the teacher-reported observations of how they viewed the children in their classrooms to be positively affected as a result of the newly implemented classroom management strategies.

Teacher outcomes. The outcomes experienced by the teachers fell into four related themes: (1) knowledge, (2) locus of control, (3) emotional well-being, and (4) teaching philosophy and practice. Broadly speaking, teachers experienced an increase in professional knowledge as a result of the course, including information about children's abilities/needs and a range of classroom management strategies. Equipped with this knowledge and the theory behind the positive strategies, the teachers increasingly experienced greater self-confidence and a wider locus of control over their ability to deal with day-to-day challenging behaviours in the classroom. This control in turn led teachers to feel calmer and less stressed in the classroom and therefore to experience improved emotional well-being in their jobs. Over time and throughout the learning process previously discussed, teachers modified their teaching philosophies to emphasise positive strategies. In this way, the four related themes are very much integrated together and a process can be traced throughout them (see Figure 4.3). While these outcomes are unlikely to have occurred exclusively in a linear progression, explanations included certain outcomes preceding others (e.g., the change in teaching philosophy and practice occurred as a result of the other outcomes and not prior to them, even though it is also possible that it reinforced the strength of the other outcomes). The remainder of this section details each of these four outcomes, illustrating how they are grounded in the data through direct quotations of the findings from the included primary studies.

The teachers reported great knowledge about many aspects related to teaching young children in schools. They particularly noted the benefit of information about children's developmental abilities and needs in understanding why children acted in challenging ways and information on how teachers could best respond to children's actions, instead of reacting in the moment.

'Teachers identified lots of examples of strategies that they learnt about either from the course content or from sharing experiences and examples with other teachers. Specific strategies included modelling behaviour, praise, coaching, labelling behaviours and feelings, ignoring negative behaviour, greeting children and parents and using rewards, amongst others.' (Marlow et al., 2015, p. 241)

'They reported increase knowledge of young children's needs and abilities in terms of a better understanding of child development, using more developmentally appropriate strategies, using more interactive teaching methods, spending individual time with children and showing children more affection. Thus teachers believed they now had more appropriate expectations for child behaviour and better understanding of the large individual differences in children's development and of the need to give attention and encouragement to all children.' (Baker-Henningham & Walker, 2009, p. 634)

'When asked how they had benefited, teachers gave two responses on average. The most common themes were: (1) having developed an increased range of effective strategies, and (2) greater use of existing positive strategies. These themes were both cited by 11 teachers (52%).' (Hutchings et al., 2007, p. 20)

By discussing the theories behind this knowledge, teachers felt more secure in this information, perhaps because they felt their understanding was went deeper. Indeed, in the one study that included qualitative findings from a long-term follow-up (Limerick), the evidence indicates that this increased knowledge as an outcome was still very much a central experience of participating teachers.

'[...] a clear increase in teacher self-confidence post-TCM training was apparent at 6-month follow-up. This improvement in confidence was also evident at 12-month follow-up, and was directly attributed to an increase in knowledge. Teachers felt better prepared in terms of having available to them, a range of evidence-based strategies for effective management of classroom difficulties.' (Hyland, 2014, p. 205)

So, not only did teachers feel that they maintained their increased knowledge over time, but this knowledge led to improved confidence. Teachers experienced greater confidence through a broadening of their locus of control, or their views on what change they could influence.

'Whilst behavioural challenges continued to present themselves from time to time, they were reported as being less frequent and teachers also felt much more confident in their ability to manage these.' (McGilloway et al., 2010, p. 19)

'Some teachers felt that the TCM course gave them "license" to change their practice and felt more confident in using the techniques and sharing them with others knowing that they could point to the theory behind them. Some teachers expressed an increase in their own self-awareness and self-management, having reflected on the way that their own behaviour directly affects children's behaviour, which made them feel calmer and more empowered to choose their response rather than react.' (Marlow et al., 2015, p. 241)

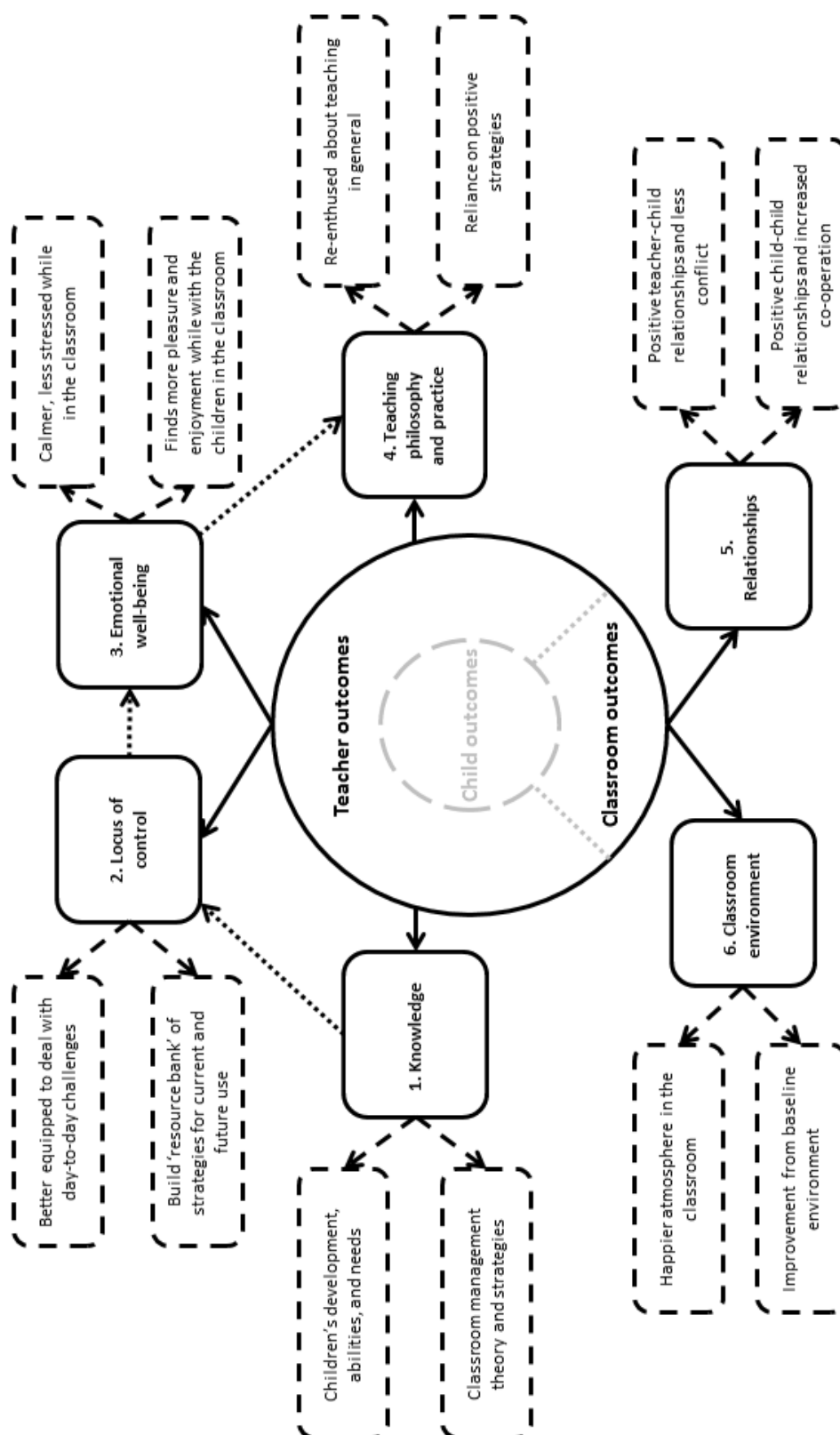


Figure 4.3 Result of the meta-synthesis: Teacher and classroom outcomes

'The use of more effective classroom management strategies led to better classroom control which made them feel less stressed, more confident and increased their ability to meet the objectives set.' (Baker-Henningham & Walker, 2009, p. 634)

'All Participants reported feeling confident or very confident that they were able to manage current behaviour problems and that they would be able to manage future behaviour problems.' (Davenport & Tansey, 2009, p. 11)

Again, this theme persisted more than a year later in the long-term follow-up interviews conducted through the Limerick study.

'Before training, teachers' confidence was generally low and they often questioned their own ability to manage their classroom and, in particular, to deal with behaviourally challenging children. Post-training findings revealed greater confidence and an enhanced ability to deal with misbehaviour as a result of the knowledge gained from participating in the training programme.' (McGilloway et al., 2012, p. 28)

'[...] teachers at baseline felt that child behaviour was largely determined by factors outside of their direct control. However, the follow-up analyses indicated that five of the six teachers had acquired a greater awareness and appreciation of their influence in this respect, thereby representing an important shift in their thinking over time. The helplessness in managing SEBD [social-emotional and behavioural difficulties] reported by teachers at baseline was not evident at 12-month follow-up.' (Hyland, 2014, p. 202)

This last excerpt, in particular, illustrates the change over time in teachers' views on the influence they possess over challenging behaviour in their classroom. It is notable that this

confidence and control of the situation extends to working with the high-risk subgroup of children who are identified with social-emotional and behavioural difficulties.

Very much tied to teachers' levels of confidence, their emotional well-being also benefited from participation in the IY TCM Programme. Teachers were able to find more enjoyment in their interactions with children. This was described as due to teachers' being able to better regulate their own emotions to remain calmer during fraught situations, and therefore experiencing less stress while on the job.

'In the current study, stress reduction was primarily due to a greater teacher awareness of how their own emotions could contribute to an escalation of low level disruptions.

Therefore, TCM techniques recommended for coping with challenging behaviour were regarded as invaluable, providing teachers (particularly NQTs [newly qualified teachers]) with the support that was absent at baseline.' (Hyland, 2014, p. 179)

'One unexpected benefit was teachers reporting increased enjoyment of the children. It appeared that the intervention had helped the teachers to relax and have fun with the children.' (Baker-Henningham & Walker, 2009, p. 634)

'[...] seven [teachers] (33%) referred to changes in themselves in terms of feeling calmer and/or less stressed.' (Hutchings et al., 2007, p. 20)

Not only did teachers feel calmer and less stressed because of changes in their own classroom, but the experience of learning and sharing stories with other teachers in the training group reduced loneliness teachers felt from being secluded within the walls of their classroom day-in and day-out.

'Teachers also reported feeling less isolated in the classroom, having met with other teachers who experienced similar challenges.' (McGilloway et al., 2012, p. 28)

Furthermore, the experiences of improved emotional well-being endured over time and were evident in the Limerick study's long-term follow-up.

'Importantly, the emotional well-being of teachers was also shown to have improved post-training. Previously, teachers reported feeling 'bogged down' and 'drained' when managing classroom misbehaviour. At 12 months, they felt considerably calmer and less stressed.' (McGilloway et al., 2012, p. 28)

Ultimately, the collection of these teacher outcomes in knowledge, locus of control, and emotional well-being can all be seen in their refreshed teaching philosophy and practice. Teachers reported experiencing a new and improved passion for teaching and reliance on positive strategies. This teaching philosophy was not an entirely new mind-set; rather, in many cases teachers expressed that they used to feel this way about teaching and were returning their daily practice back to their ideal approach to teaching. However, the experience of being on the IY TCM training course and observing the positive changes in their classrooms led teachers to reset their professional attitude and focus with renewed gusto on delivering a true-to-their-teaching-philosophy classroom experience for the children each day.

'[...] the constant pressure of classroom management had exerted a significant toll on the emotional well-being of most teachers and, in some cases, caused some degree of professional incongruity. Some reported acting in ways that contradicted their personal teaching philosophy; they felt that they were becoming the kind of teacher they never wanted to be. [...] Encouragingly, the follow-up analysis illustrated a significant shift in

the outlook of four teachers, who felt that TCM training had imbued in them a greater enthusiasm for their role as educators.’ (Hyland, 2014, p. 181)

‘[...] teachers found that improved classroom interactions (e.g., the use of praise and encouragement) had a powerful impact, not only on the children but also with regard to how they themselves felt. They reported an energising effect, feelings of optimism towards the future, better preparedness for incoming classes, and a renewed professional enthusiasm for teaching.’ (Hyland, 2014, p. 181)

‘Teachers gave many clear examples of how they used the strategies taught to prevent and reduce child misbehaviour in the classroom and their accounts revealed that increases in the use of positive strategies resulted in decreased reliance on corporal punishment and other negative strategies such as shouting and name-calling.’ (Baker-Henningham & Walker, 2009, p. 634)

‘Nine teachers (43%) referred to a change in their philosophy [of teaching].’ (Hutchings et al., 2007, p. 20)

As with the other teacher outcomes, the change in teaching philosophy continued through long-term follow-up.

‘Continued implementation of the TCM programme also resulted in a reduced need to use strategies such as ignoring, or time-out. Consequently, at the beginning of a new academic year, teachers were able to focus on further implementing and developing an enhanced classroom environment.’ (McGilloway et al., 2012, p. 27)

‘Two teachers specifically used the term ‘second nature’ to describe the extent to which they had successfully integrated the TCM philosophy into their classroom management.’ (Hyland, 2014, p. 200)

In addition to these teacher outcomes, which are experiences directly felt by participating teachers, there were classroom outcomes, which involved the teachers as well as the collective class as a unit.

Classroom outcomes. Classroom outcomes refer to the whole-group outcomes of the teacher and children. Two themes emerged in this area: (1) relationships and (2) classroom environment. Reports of improved teacher-child and child-child relationships, at the classroom level as opposed to an individual (target child) level reflect the move by the teachers to focus on the positives exhibited in the classroom rather than the negatives

‘Prior to training, many classrooms were tense environments where the learning opportunities of children were limited by ongoing disruption. Verbal (and in some cases, physical) child aggression posed clear difficulties for teachers. The post-TCM class environment was reported by all teachers to be a very different place, where children felt secure and where teachers and students enjoyed positive relationships. Not only did teachers note a lower incidence of conflict with children, they reported improvement in peer relations.’ (Hyland, 2014, p. 185)

‘Irrespective of whether their class included children with behaviour difficulties, 20 teachers (95%) reported a general benefit to their class as a whole and, of the 12 who reported having children with specific behaviour difficulties, 11 (91%) reported that the

training had helped them to become better equipped to deal with and teach these challenging children.’ (Hutchings et al., 2007, p. 20)

‘It was clear from the findings, that the teachers’ views of the classroom prior to training [...] had been transformed once they had completed the course. The general atmosphere within the classroom was considered to be calmer and more co-operative, whilst it was also regarded as a happier and more positive environment in which to work (and learn).’ (McGilloway et al., 2010, p. 19)

With the broadening of this sphere of encouragement and co-operation to the level of the classroom, teachers also experienced improved relationships with the parents of their pupils.

‘Some of the teachers also alluded to the positive impact of the TCM programme on home-school communication and collaboration. For example, some participants felt it was rewarding to pass on information to the parents about the programme (e.g. how it is working) whilst others indicated that, as a result of the programme, they were making more effort to contact and share information with parents about their child’s progress.’ (McGilloway et al., 2010, p. 18)

‘This increased focus on positive rather than negative behaviours also resulted in teachers’ developing better relationships with the children and with their parents.’ (Baker-Henningham & Walker, 2009, p. 634)

As with all of the other experienced outcomes thus far, long-term follow-up data suggests that the improved relationships and warmer classroom environment continued with the introduction of a new class of pupils in the subsequent academic year.

‘At 12-month follow-up, they reported that the children wanted to learn and co-operate, leaving them free to teach from the distraction of ongoing behavioural challenges and classroom conflict.’ (McGilloway et al., 2012, p. 27)

‘One year after training had begun, classrooms were regarded by intervention group teachers as calm, warm, positive environments, where children wanted to learn and achieve.’ (Hyland, 2014, p. 208)

For all of the outcomes at the teacher and classroom level, teachers also reported observing a number of outcomes at the child level, which are detailed in the next section.

Child outcomes. The stakeholders (predominantly teachers) also reported the experience of observing positive outcomes at the child level. Not surprisingly given the IY TCM programme’s focus on (1) reducing problematic behaviour and (2) increasing pro-social behaviour, these two areas received the most attention in the included studies. Also addressed, however, were children’s (3) emotional well-being, and, less so, (4) academics, and (5) other environments (see Figure 4.4).

Teachers experienced reductions in children’s challenging behaviours such as inattention, disruption, and aggression, and these outcomes were directly attributed to the use of IY TCM-learned classroom management strategies. They felt that children were able to stay focused and were generally more appropriately behaved while in the classroom. Not only did the teachers experience this outcome, but the head teachers also noticed receiving fewer child referrals for discipline.

‘Several changes in child behaviour were reported by teachers at the post-intervention stage. Most notably, teachers reported that children were displaying a greater ability to stay on task, were less distracted, and were engaging in less attention-seeking behaviours. Children who, at baseline, had posed significant challenges to teachers in terms of their attention-seeking and lack of task-focus, were seen as much more manageable in the class. This was regarded as highly significant by teachers and due directly to the implementation of the TCM principles.’ (McGilloway et al., 2010, p. 18)

‘Some teachers commented on having seen the positive effect of strategies within their classes, for example changes in children’s behaviours, and that this had been noticed by others in their schools such as heads.’ (Marlow et al., 2015, p. 241)

‘Teachers reported reductions in the problem behaviour of children who had challenging behaviour at baseline, improvements in the behaviour of shy children and reductions in the levels of aggression and noise in the classroom.’ (Baker-Henningham & Walker, 2009, p. 634)

‘All teachers said that pupils had benefited from their training and gave an average of three responses to this question. The most common response – from 14 teachers (67%) – was that the strategies had helped children to pay more attention [...]’ (Hutchings et al., 2007, p. 20)

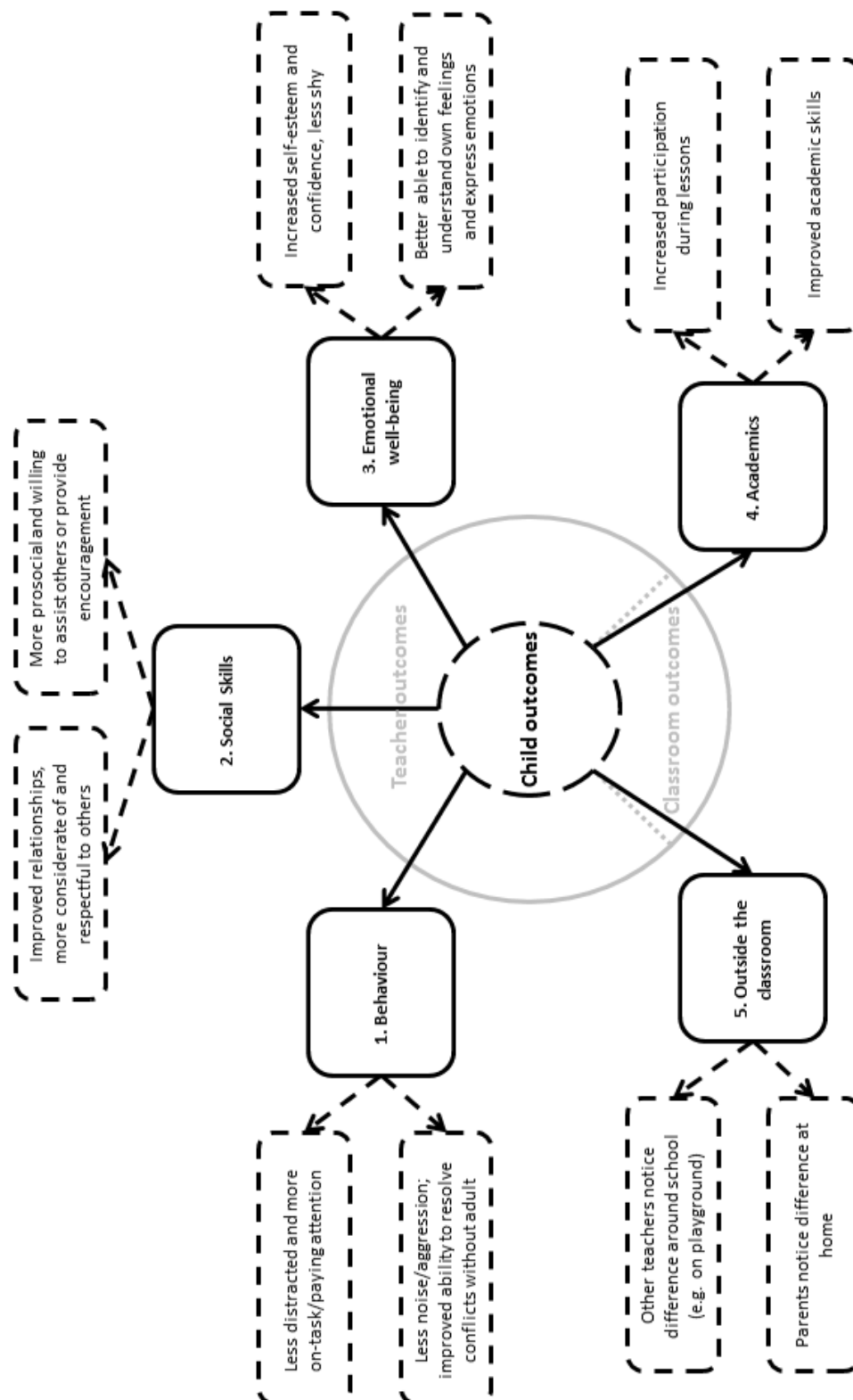


Figure 4.4 Result of the meta-synthesis: Child outcomes

'All [teachers] reported an improvement in the behaviour of individual students on whom specific guided interventions had been targeted.' (Davenport & Tansey, 2009, p. 11)

Not only did problematic behaviour decrease, but positive behaviours increased.

Specifically, children's social skills were seen to have benefited from the strategies teachers had learned on the IY TCM training and were now implementing in their classrooms. Teachers reported that the children co-operated better with each other and were able to problem-solve among themselves when conflict arose, rather than require the teacher's intervention to quell arguments.

'Another notable aspect of child behaviour change, reported by teachers post-intervention, was the increased ability of their pupils to independently resolve conflict with their peers. Prior to the TCM, teachers were often called upon by the children to intervene in the conflict, but this was a much less frequent occurrence in the post-TCM period. One teacher reported that this ability to resolve conflict with peers, extended outside the school environment. Children were seen to have developed these skills and used them in a variety of settings by labelling emotions and talking through problems in the classroom context.' (McGilloway et al., 2010, p. 18)

'Teachers also reported that children had shown improvements in their overall conduct within the classroom. Several different aspects of child behaviour were discussed by teachers as having noticeably changed from pre- to post-TCM. Children had developed better peer relations, they had become more pro-social, and showed a greater willingness to help and encourage each other. The modelling of positive behaviour by teachers was regarded as being instrumental in promoting pro-social behaviour in the class, even

amongst children who had previously displayed good class behaviour.’ (McGilloway et al., 2010, p. 19)

‘Teachers reported improvements to children’s social skills, including increases in sharing, helping and complimenting each other and it was clear that these skills were highly valued.’ (Baker-Henningham & Walker, 2009, p. 635)

‘[Teachers reported that children benefited] by becoming more considerate of others (11 teachers, 52%) [...] Other responses included seeing children become more respectful (eight teachers, 38%), and children enjoying responsibility (six teachers, 29%). This latter response is further evidence of how the collaborative course style filters through to the classroom, with teachers finding ways to give greater responsibility to children.’ (Hutchings et al., 2007, p. 20)

‘Two teachers in the current study also regarded the TCM programme as effective in developing the social skills of non-disruptive children, an issue which has received little attention in previous research. In fact, the relationship building features of the TCM programme were regarded by two of the interviewees as having had direct, tangible benefits (both in school and at home) for children who were more introverted.’ (Hyland, 2014, p. 204)

These outcomes were experienced not only among children who were considered to be lacking in such skills at baseline, but also by children who would not have been targeted

by the teacher as problematic. This particular finding is notable because it suggests that all children can benefit from these positive management strategies.

Furthermore, not only had children's social skills grown, but teachers felt that the children's emotional well-being had been boosted as a result of the training the teachers had received. Children demonstrated to their teachers greater abilities in recognising how they were feeling, naming the emotions, and responding appropriately in various emotional states.

'Children were thought to have benefitted from the labelling of emotions. This was a key aspect of the TCM programme which teachers believed was extremely beneficial for the children in their class. Participants stated that, prior to the TCM programme, children had been unable to identify the emotions that they experienced and felt that this had led to inappropriate behaviour in the classroom. Post-training, teachers were able to model appropriate responses to emotions; children learned from this and were thereby equipped with the language to express their feelings. This, in itself, was perceived by teachers as increasing the confidence of children, enabling them to speak their minds in the class, rather than 'acting out' in frustration.' (McGilloway et al., 2010, p. 18)

'Teachers also reported benefits to children's emotional competence including better understanding of their emotions, showing more empathy, better anger management skills and showing increased confidence.' (Baker-Henningham & Walker, 2009, p. 635)

'[Teachers reported that the training they received resulted in] boosting children's confidence (10 teachers, 48%) [...]' (Hutchings et al., 2007, p. 20)

While only one study touched on children's outcomes in academics and how all of the outcomes generalised to other environments, the findings are included in this meta-synthesis because they continue the line of reasoning that flows through this synthesis and do not contradict findings from the other studies. Teachers experienced children's increased engagement in school and felt that their academic skills had improved as a result of the implementation of IY TCM strategies. Teachers were reinforced in their observations of children's benefits from the positive strategies from conversation with parents, who said the children's outcomes were apparent at home, too.

'Other benefits reported included increase participation in lessons, better academic skills and increased enjoyment of school. The benefits to children's participation and learning were attributed to the interactive teaching methods teachers were now using.' (Baker-Henningham & Walker, 2009, p. 635)

'Some teachers reported that there was transfer of the skills learned to the home setting and it was interesting that several teachers reported improvements to the relationships between parents and their children.' (Baker-Henningham & Walker, 2009, p. 635)

Collectively, these included studies illustrate an intricate network of experiences as a result of the IY TCM programme and the implementation of IY TCM strategies in classrooms. Although most of the stakeholders included in these studies were the teachers who had participated in the training, the little data reported from facilitators and head teachers as well as secondhand data about parents' experiences supports the conclusions drawn. Not only did teachers undergo a complex and iterative learning process throughout the months of training and practice, but they experienced the fruits of their labour in the many reported outcomes for themselves, their entire classroom, and the children within their class.

Critiques of the programme. Despite the many positive experiences shared by the various stakeholders, critiques of the IY TCM programme were also present in the literature. The critiques predominantly appeared in sections about recommendations for future implementation, barriers to implementation, and difficulties with the programme. Unlike the previous discussions on the cycle of learning and perceived outcomes, the critiques were less cohesive across studies, and it appears that there was greater variability within and across studies on whether certain aspects were considered negatively by a participant or not.

Despite this relative dearth in data, a preliminary model of critiques of the programme is presented here. The model once again uses a cyclical process of recruitment, delivery/content, outcomes, and scale-up to illustrate not only the many times throughout the process that critiques are raised, but also the difficulty of balancing competing aspects of programme delivery. Therefore, this model is not meant to provide a one-size-fits-all answer on how to address these critiques, but rather to situate the critiques in the process so that decision-makers can ascertain how these might resonate in their own context and consider appropriate responses to their specific situation (see Figure 4.5).

Reason for taking the course. Motivation for participating on the IY TCM programme was highlighted as an important factor, with the distinction between teacher choice to participate contrasted against compulsion. In the Limerick study, the group facilitators expressed that teachers who willingly participated on the IY TCM course would be more open to the training compared to those who were informed that they had to participate. The suggestion is that teacher choice could be a moderator for teacher engagement in the course and its material, such that those teachers who wanted to take the course would experience positive effects at its conclusion while those who did not want to take the course but who had to do so would experience no effects or even negative effects.

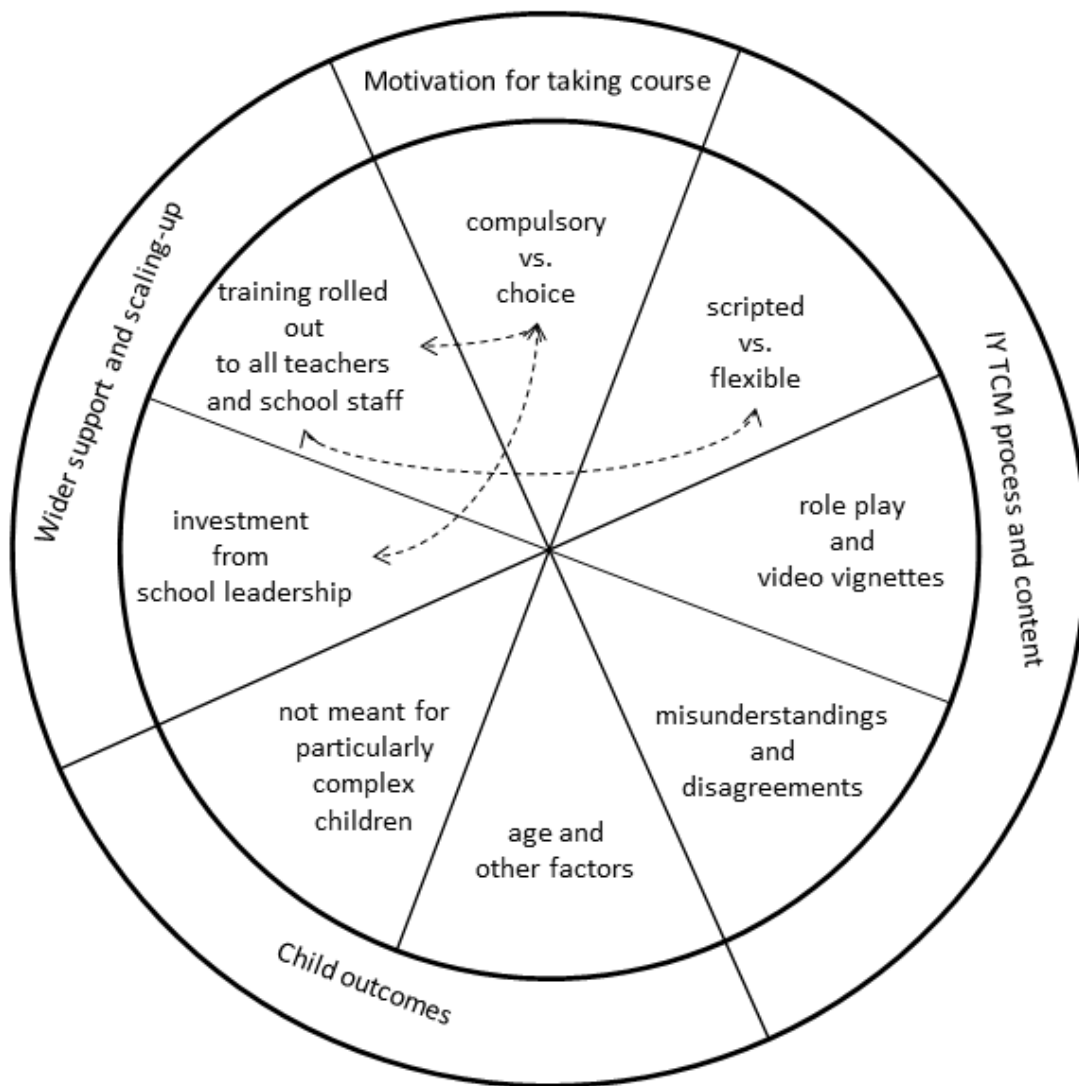


Figure 4.5 Result of the meta-synthesis: Critiques of the programme

‘Both facilitators in the current research also believed it to be imperative that teachers themselves choose to attend TCM training; compulsory attendance of teachers to the programme (i.e., mandated by school management) is a poor foundation for learning. They made reference to the difficulty of previously trained groups who were ‘forced’ into attendance by well-intentioned school principals. For example, one facilitator stated how she had delivered many training sessions in the past where attendance had been

compulsory, and attendees as a result had antagonistic attitudes towards participation and were less willing to engage with the learning process.’ (Hyland, 2014, p. 194)

The Devon study also reported the variety of ways by which teachers were recruited to participate on the IY TCM course, although in this study the researchers did not observe cause for concern that teacher choice (or lack thereof) resulted in different experiences.

The caveat here is that the researchers note their inability to incorporate the experiences of teachers who discontinued the training programme, leaving a gap in our knowledge of how this group experienced the programme.

‘Teachers had different experiences of the way in which they were selected from within their school to attend TCM. In some schools, a group of teachers had been offered the opportunity and individuals had put themselves forward; in other schools particular teachers had been asked to attend by the head for varying reasons. There did not appear to be a link between the way that teachers were recruited and their experience of the course (although we were unable to collect detailed views of the teachers about this selection process or those that left the course).’ (Marlow et al., 2015, p. 243)

Certainly, the influence of teacher choice in participating on the training programme and any potential change in effectiveness has very real implications for how decision-makers move forward when scaling up implementation of the IY TCM programme at different levels. However, the qualitative evidence available within this meta-synthesis is so minimal that beyond confirming that teachers had varying levels of say in whether or not they sat the course, the contradictory information on whether this might influence experiences and/or outcomes does not lead neatly to one conclusion.

IY TCM process and content. Among teachers who participated in the course, there were three main parts of the process of the IY TCM programme that emerged in their

critiques: its manualised nature, specific learning techniques, and the presence of extra challenges. The group facilitators in the Limerick study expressed strong views in favour of implementing the IY TCM course in accordance to the manual. This was not due to their individual preferences for each step of the programme but rather linked to desired outcomes, such that implementation fidelity was suggested as a possible mediator of outcomes, both initially and at follow-up.

‘Within the current study, the issue of implementation fidelity was emphasised throughout by facilitators, and the standardised delivery of TCM training was seen as integral to successful short- and long-term outcomes.’ (Hyland, 2014, p. 193)

In contrast, views from participating teachers illustrate a desire for more flexibility from the designated schedule. In particular, teachers expressed some dissatisfaction with what they felt was an inflexible curriculum that did not adequately adapt to meet their levels of experience and resulted in unnecessary repetition of concepts and unfavourably long training days.

‘Although the programme was received very well, all teachers highlighted aspects of it with which they were less than satisfied, such as the perceived repetition of topic information. The general consensus was that the material could have been covered as thoroughly without the need for constant repetition throughout the day. The intervention group teachers considered the length of the training day (i.e. 9.00am- 4.00pm), in particular, to be excessive.’ (McGilloway et al., 2010, p. 19)

‘Some teachers expressed a view that the level of delivery was sometimes pitched too low and that they felt there was a “script” that the leaders had to follow which could have been more flexible. However, some thought that the leaders got the level of delivery just

right and others acknowledged the need for balance to meet all needs.’ (Marlow et al., 2015, p. 240)

Although we can see from this last excerpt that the teachers recognised the spectrum of differing needs within the group, it is apparent that the fidelity-adaptation discourse is very much relevant to how this programme might be taken up. Particularly given the previous section’s discussion of the role of choice in teacher engagement with the training, a pertinent question remains of whether an inflexible programme can engage teachers across the spectrum of experience and whether this engagement (or lack thereof) would subsequently influence outcomes. Furthermore, if flexibility was introduced, how would it be incorporated so as to avoid negatively influencing outcomes by skipping over core components or reducing the dose of important elements? This is as yet unclear, but an area requiring attention based on these reports.

When it came to specific learning techniques used in the process of the IY TCM course, two techniques stood out as eliciting very strong opinions among participating teachers: role plays and video vignettes. There appeared to be general reluctance to engage in role play as a form of practicing the IY TCM strategies, perhaps in part due to shyness or embarrassment and in at least one case due to feeling patronised by the activity. The discomfort felt by participants when faced with role playing tasks highlights the importance of an earlier theme from the cycle of learning: establishing the group dynamic and creating a safe space. Indeed, despite the strong reservations towards role play, this technique was also identified as a valuable part of the learning experience.

‘There was also significant variation in teachers’ beliefs regarding TCM learning tools. The use of role play was a particularly divisive issue. For example, one teacher felt patronised when she engaged in this activity [...] Two of the teachers highlighted

reluctance among their peers to engage in role play and that this had negatively impacted its overall effectiveness. Both facilitators noted that despite initial teacher reluctance to participate, all eventually engaged with this learning tool, and noted its utility in experiential learning. In fact, one teacher reported that, although the group tended to shy away from role play, it was this form of active participation which consolidated her own learning.' (Hyland, 2014, p. 190)

'Teachers talked about the challenges of practising skills within the sessions (e.g. role play, coaching, puppet work), but also placed great value on the process of seeing the practice [...]' (Marlow et al., 2015, p. 240)

Unlike critiques of role playing techniques in the learning process, criticisms of the video vignettes centred primarily on cultural differences. In all four studies where the video vignettes were criticised, participants noted that the American classrooms portrayed in the vignettes were incongruent with their own settings. Even though it was acknowledged that the vignettes could still serve as conversation starters for IY TCM topics, the views expressed by teachers raise important questions regarding not only teacher engagement during training sessions, as previously mentioned, but also acceptance of the programme.

'Similarly, the use of vignettes elicited divergent opinions [...] although some teachers viewed their use as an integral part of the learning process (e.g., allowing a greater understanding of the mechanisms for dealing with behavioural problems), others did not. Two teachers criticised very different aspects of vignette usage. One believed them to be over-used whilst the other (despite reporting that she enjoyed watching them and recognising their benefit) found the American accent and demeanour to be "a bit OTT" (T8).' (Hyland, 2014, p. 191)

'Teachers gave lots of feedback on the (American) video-clips; many felt they did not fit the English primary school culture and were dated. Some felt that the videos nevertheless had a value in encouraging discussion and seeing the strategies work in "real life" situations.' (Marlow et al., 2015, p. 241)

'Some teachers also recommended that some video vignettes of Jamaican classrooms should be used rather than only clips from US classrooms which are so different in terms of staffing, classroom resources, class size and space.' (Baker-Henningham & Walker, 2009, p. 637)

'Some expressed a [...] desire to see video vignettes of children in Irish school settings.' (Davenport & Tansey, 2009, p. 12)

The Kingston (1) study reported on the challenge of dealing with teachers' misinterpretation how/when to apply IY TCM strategies in the classroom as well as disagreeing with core tenets of the IY TCM approach. Although other studies referenced the importance of time to rehearse and imbed strategies into teachers' day-to-day practice, none raised concerns about serious confusion or fundamentally incompatible approaches to classroom management, as were found in the Jamaican context.

'Several teachers also reported having difficulty implementing strategies to deal with child misbehaviour (ignore and time out) correctly and they found that some of the new strategies required practice and effort before they were confident implementing them. There was evidence that the use of the discipline strategies had not been fully understood

by several teachers. For example, the ignore strategy is meant to be used for minor child misbehaviours that are not causing harm to others and that do not involve non-compliance to teacher requests. However, one teacher reported that it was difficult to ignore non-compliance while another teacher described how she could not ignore aggressive behaviour. One teacher commented that although the strategies were useful, there was still a place for corporal punishment for some of the more difficult children.’ (Baker-Henningham & Walker, 2009, p. 636)

It is here, perhaps, that contextual information might provide some insights into why the main difficulties with implementing the IY TCM programme in Jamaica were not present in any of the other studies. It is of particular note that in the Kingston (1) study, teachers were qualified at the paraprofessional (i.e., teaching assistant) level, which is in contrast to the further preparation and certification that would have been present among teachers in all of the other included study contexts (Ireland, England, and Wales). This discrepancy in prior professional training could explain in part why some teachers in Jamaica struggled to appropriately apply learned IY TCM strategies such as ignore, over and above needing time to practice. Furthermore, the study authors note that ‘Different cultural beliefs and values may also affect the acceptability of the goals and/or strategies advocated by these programmes’ (Baker-Henningham & Walker, 2009, p. 633). This is particularly relevant for the disagreement over the use of corporal punishment as a strategy used by teachers in their classrooms. It is possible, that in other contexts where teachers have lower baseline levels of teacher training in classroom management or where there is a culture of corporal punishment, extra work might be needed to ensure that the IY TCM programme is acceptable and accessible to participants. The other side of this issue is that these contexts

might present opportunities for the most growth and therefore largest effect sizes on outcomes, if teachers take on board the strategies.

Child outcomes. Previously, this synthesis explicated the many positive child outcomes that key stakeholders reported observing or experiencing. Despite the optimistic and encouraging litany of benefits, there are still important critiques to note. Teachers mentioned outside factors such as maturation and strategies already in place, suggesting that these – alone or with the combination IY TCM strategies – could explain outcomes.

‘While perceptions regarding the effectiveness of the overall TCM programme were very positive (n = 8), two teachers noted that changes in child behaviour may have been influenced by non-IY TCM factors. For example, one teacher stated that by the time post-intervention data were collected, the children were a little older and positive class outcomes may have been due to a combination of the TCM programme and child maturation. Another teacher suggested that the positive changes in a child’s behaviour were partially attributable to her efforts prior to TCM programme training.’ (Hyland, 2014, p. 185)

Other criticisms mentioned a relative lack of attention on how to implement the IY TCM strategies with older children compared to younger ones, suggesting that an area further programme development could involve specific attention directed towards lower primary versus upper primary school classrooms.

‘Some teachers felt that some of the course content was targeted at younger pupils (aged five to seven) and would have liked more examples with older year groups.’ (Marlow et al., 2015, p. 241)

Finally, some critiques levelled at the effectiveness of the IY TCM programme identified subgroups of children whose needs were so great or different that they would not benefit from the application of these strategies alone.

‘Generally, teachers regard the IYTCM strategies to be of practical use in improving classroom discipline, although it was noted that for a minority of children, some techniques were ineffective. For example, the inclusion of children with special needs in the mainstream classroom and class size/age of pupils were seen as a challenge to implementation. Overall, teachers recognised that the programme worked, but not for every child.’ (McGilloway et al., 2012, p. 30)

‘[...] any unsuccessful outcomes were thought to result more from the complexity of child need, than from the inadequacy of the TCM programme. For example, one teacher implemented TCM strategies in her classroom while awaiting psychological assessment for a child. Whilst the strategies did not work on the target child, they were effective for other children in the class.’ (Hyland, 2014, p. 184)

The experience that particularly complex children might not respond to the IY TCM strategies leaves some teachers in a quandary. The IY TCM programme was originally designed to address children with clinical and pre-clinical problematic behaviours in school, suggesting that it is meant for high-risk children. However, ‘psychological assessment’ could encompass any number of mental health needs outside of the realm of conduct disorder, and it is very possible that complex or high-risk children who do not respond to the IY TCM strategies require interventions targeting different underlying causes. If this is the case, then the issue is not ineffectiveness of IY TCM but of inappropriate intervention to address a problem. Furthermore, it is important to note that

programmes demonstrating effectiveness at the sample level are assessed by the difference in means, and do not represent individual trajectories or even suggest that all children will benefit. If subgroups of children are systematically not benefiting from application of IY TCM strategies, it could suggest that they require additional or different supports. The evidence available in this synthesis is insufficient to draw any conclusions regarding whether subgroups of children (e.g., those with special educational needs, in upper primary school) are indeed systematically underserved with IY TCM techniques. Nevertheless, there is preliminary experiential evidence that the effects of the IY TCM programme might be moderated by children with complex or special educational needs and those who are older.

Wider support and scaling-up. Finally, after recruiting participants, delivering the programme, and observing any effects, key stakeholders expressed their views on the role of wider support for training and scaling-up implementation beyond each study's limited remit. This theme is incorporated under the critiques of the programme synthesis because it illustrates the constraints present in the real-world, which could moderate programme effectiveness. Additionally, it demonstrates a need to consider the level of implementation and contributes to the theory that intervention might need to occur at an additional, higher level of cluster: the school rather than the classroom.

Firstly, the influential role of school leaders was noted. IY TCM facilitators in the Limerick study reported that the heightened interest of school heads was at least in part responsible for teacher motivation.

'The level of teacher programme engagement in the current study may be related, at least in part, to the support of principals, who along with initiating school participation, also built considerable enthusiasm among staff regarding potential TCM outcomes. [...] The

facilitators noted that the teachers in the current study were more motivated to learn than some previous cohorts whom they had trained [...].’ (Hyland, 2014, pp. 194-195)

More than the support of school leaders to back individual teachers participating in the IY TCM course, a common and prominent theme emerged from the included studies of the potential future benefits of widening the scope of the training to the whole school staff and even to parents.

‘Some teachers also discussed the benefits of adopting a whole-school approach in order to maximise the effectiveness of the programme and of involving more parents in the new strategies to better manage challenging child behaviours. One teacher feared that without a whole-school TCM implementation, the improvements gained in child behaviour may be short-lived.’ (McGilloway et al., 2010, p. 19)

‘Some teachers talked about sharing TCM strategies within their school, either in their own classroom (e.g. with teaching assistants) or at staff meetings. Most dissemination happened on an informal and ongoing or “as required” basis, and often involved teachers explaining the approaches, but also included teachers modelling strategies to colleagues through their own practice. Some teachers talked about the greater effectiveness of TCM strategies when they are consistently applied (e.g. by teaching assistants, other teachers, heads) and that this can be a challenge.’ (Marlow et al., 2015, pp. 241-242)

‘The main recommendations from the teachers were that the skills they learnt were so valuable that all pre-school and primary teachers should attend the training [...] Teachers suggested that they should share the strategies learned with parents and with other

teachers. For example, some teachers expressed significant concern that other schools were in need of the intervention and they reported actively helping them.’ (Baker-Henningham & Walker, 2009, pp. 636-637)

‘[...] two teachers (10%) reported that the inconsistency of other staff was a barrier [...] There was a strong feeling among the teachers that other, non-teaching, staff in the school also needed training in behaviour management, with seven (33%) mentioning teaching assistants and others mentioning head teachers, playground and canteen staff.’ (Hutchings et al., 2007, pp. 20-21)

Despite these studies taking place across different countries and cultures, which produced contradictory critiques and recommendations in previous themes, the consistent message of widening the training to include all teachers and staff suggests that the theory of change needs to be expanded to the school level, rather than classroom level. With inconsistencies in other staff’s management strategies hindering the effectiveness of the IY TCM programme on child outcomes according to some stakeholders, we see the importance that teachers placed on the school as a whole community, rather than just their classroom as isolated sub-communities within schools. If, indeed, the involvement of school leaders and the implementation of the IY TCM programme at the whole school level is a better approach to intervention, then it is entirely possible that results from effectiveness studies may well underestimate the effect sizes because the programme has not been comprehensively rolled-out. Of course, it is possible to expand yet further beyond the school to a region or even the national levels.

‘Unsurprisingly, facilitators were also invested in the concept of a national roll-out of the TCM programme. For instance, one mentioned the possibility of including the programme

in Department of Education in-service training. However, she noted [...] that taking elements of the training programme out of their intended context could be problematic.' (Hyland, 2014, p. 193)

Such wide-scale take-up of any training brings the conversation back around to the original critiques of the programme: reason for taking the course. As already presented, those teachers who had a choice to participate on the training course quite possibly were more engaged in the process and open-minded to the content compared to those teachers who were forced to attend by their school head. When the suggestion arises, as it has, that all teachers and school staff should have access to this course because it was so well received by study participants, this begs the question of how to do so without making it compulsory and risking engagement. It is here, in particular, that research is needed to determine whether in fact teacher choice moderates teacher or child outcomes, so that we can advise decision-makers on how important it is to involve choice when offering this training to staff.

Discussion

Summary of main findings. The findings from this qualitative systematic review and meta-synthesis of stakeholders' experiences suggest that the IY TCM programme is acceptable and enjoyable for participants. The cyclical process of engaging in the learning process and materials individually as well as collectively produced a pattern of learning, hands-on practice, reflection, adjustments, and further practice. Along the way, teachers felt as though they were building up a resource bank of strategies for future use.

Stakeholders definitely experienced numerous outcomes attributed to the training programme and across teacher, classroom, and child levels. Despite the many positive views of the IY TCM programme, there were a number of critiques levelled against the process and content. Most notably are the reservations to engage in role play, use of

culturally discordant video vignettes, and necessity of scaling-up training to the school level at minimum.

Quality of evidence. The quality of evidence from the included studies was variable, and the largest barrier was a general lack of depth across all included studies. While the Limerick study consisted of multiple records including a PhD dissertation, resulting in a relatively rich description of the primary findings, other studies (e.g., Dublin, North Wales (1)) were extremely brief in their reporting and relied heavily on frequency counts and quasi-statistic descriptors (e.g., all, many, some) to present the data.

Potential biases in the review process. The most notable limitation of this review and qualitative meta-synthesis is the small number of included studies and the limited depth and quality among those few studies. Where contradictions existed within the synthesis (e.g., critiques section), these were highlighted and discussed. Often, however, studies did not contradict each other but instead bias potentially crept in due to the predominant role the Limerick study played in the synthesis (it served as the index paper because of its high quality and rich descriptions). Although the Devon study was second in its contributions to the overall meta-synthesis, in no way should these findings be considered to have reached saturation. On the contrary, the most robust sections of the synthesis, are determined by convergence on the themes from multiple included studies. All of the teacher outcomes (knowledge, locus of control, emotional well-being, and teaching philosophy) were grounded in data that spanned the included studies. Similarly, the child outcomes of behaviour, social skills, and emotional well-being were robustly justified in the data across multiple included studies. The themes of role play/video vignettes and scaling-up implementation to the whole school emerged from data across the included studies, too. The classroom outcomes, remaining child outcomes, and other

critiques are not as strongly substantiated in the available literature and should therefore be interpreted with caution.

Despite these limitations, the qualitative meta-synthesis remains a valuable contribution because it is the first to examine what the available experiential data on the IY TCM programme is and how new theory can emerge from understanding the collective findings from each primary study.

Conclusion

Implications for practice and policy. The results from this systematic review and qualitative meta-synthesis of stakeholders' experiences of the IY TCM programme illustrate the complex learning process that teachers undergo and the benefits of this training as well as some critiques. Synthesising this information adds valuable contextual information to the evidence-base for decision-makers, who can make more educated choices regarding whether the teachers in their school or area might benefit from embarking on such a journey and if so, how the training will be introduced, implemented, and measured.

Implications for research. The lack of rich descriptions was apparent, despite the common practice of conducting qualitative studies prior to or parallel with trials of effectiveness. The rigour of the qualitative studies did not match that of the quantitative ones, and future research in this area should seek to balance this discrepancy. A number of possible moderators and mediators were identified based on stakeholder experiences and explanations of how they viewed the IY TCM programme worked (or didn't) for them. These potential mechanisms should be the focus of future research, to provide deeper insights into how this training course works, for whom it is more effective, in which conditions, and when.

Chapter 5 : Cross-synthesis of evidence from trial effectiveness and stakeholders' experiences

Results

This chapter presents the results from the cross-synthesis of the findings from both the quantitative strand on trial effectiveness (Chapter 3) and the qualitative strand on stakeholders' experiences (Chapter 4). The overarching themes from each strand were thematically analysed using the framework approach and visually organised in a series of integrative grids. Grids were drawn two different ways. First, the quantitative findings filled the column headings and the qualitative studies were listed down the rows. Findings from the qualitative strand populated the cells. Second, the qualitative findings occupied the column headings, with the quantitative studies inserted into the rows. Findings from the quantitative strand were inserted into the cells: first at the multilevel meta-analysis level and then any additional quantitative findings from the strand that were not meta-analysed, each added to the appropriate study row. For practical purposes due to the scope of the qualitative findings, this process is illustrated across two grids: one with the cycle of learning and outcomes constructs and one with the critiques constructs. Approaching the cross-synthesis from this perspective meant that neither strand was valued as the 'starting point', with the other strand needing to fit within the first's framework. It also meant that instances where one strand explored themes not covered by the other strand, these data were not relegated to a catchall column without exploration from the other perspective.

In addition to the integrative grids, the findings are more succinctly summarised throughout this chapter in a series of figures that visually illustrate: 1) whether the quantitative and qualitative findings converge (green) or diverge (red); 2) the direction of the findings (up arrow indicates increase, down means decrease); and 3) the extent to

which we can be confident in these results based on the coverage of each topic across the included studies (darker shades indicate strong coverage, lighter shades indicate weak coverage).

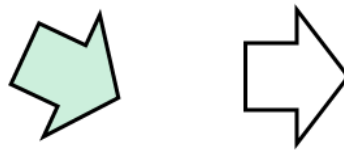


Figure 5.1 Example of visual illustrations of cross-synthesis findings

For example, the left arrow above (Figure 5.1) stands for an outcome in which the quantitative and qualitative findings are in agreement (green), the outcome of interest decreases as a result of intervention (negative slope), but which is only covered in few studies (light shade). In cases where one strand does not address a construct found in the opposite strand, a blank neutral arrow points to the right, indicating no (dis)agreement between strands by its absence of colour, no direction of findings by its zero slope, and little confidence in any conclusions based on lack of coverage by its light shading.

Overview of findings

Broadly speaking, this cross-synthesis concludes that areas commonly covered across both quantitative and qualitative strands produced harmonious findings: with data from each strand suggesting similar trends or results. Generally, the quantitative evidence supports conclusions that IY TCM intervention is more effective than standard practice across studies, and qualitative teacher reports from these studies confirm that teachers feel their practice and children's behaviours have improved. This finding is encouraging in that decision-makers can rest assured that teachers and researchers alike are apt to welcome this intervention.

The novel findings from this cross-synthesis lie predominantly in the areas that are not covered (or only partially addressed) by both strands. This dissonance illuminates different research priorities and ensuing gaps in our understanding of the processes and results of implementing the IY TCM programme (see Table 5.1). Notably, the quantitative evidence has lacked a focus on revealing and evaluating the possible mechanisms of change by which the intervention achieves its intended outcomes. Meanwhile, the qualitative evidence is missing input from most of the indirect key stakeholders (e.g., parents, children), relying almost entirely on views from teachers who have successfully completed (and not dropped out of) the training course. Furthermore, while in some cases the qualitative research was conducted as part of feasibility or piloting work to inform a larger quantitative trial, the dearth of quantitative data on many of the themes that emerged from the qualitative data suggests that the integration of evidence across the quantitative-qualitative divide has been limited, and future work should seek to more fully and continuously weave together knowledge and experience gathered in both numerical and non-numerical forms.

Mapping the qualitative findings onto the quantitative constructs. Although the qualitative studies included in this review generally lacked depth in their reports, as evidenced from the first grid (see Figure 5.2), three out of the five studies (Limerick, Kingston (1) , North Wales (1)) provided very good breadth on issues relevant in the quantitative studies. One study (Devon) examined feasibility issues, which is perhaps partially why its breadth appears limited when mapped onto the quantitative constructs. One study (Dublin) was reported so sparsely that its narrow overlap with the quantitative constructs was expected. Findings are reported in four categories: 1) teacher outcomes, 2) classroom outcomes, 3) child outcomes, and 4) parent outcomes (see Figure 5.2 for summary and Figure 5.3 for integrative grid).

Table 5.1 Themes from quantitative and qualitative strands

	Quantitative strand	Qualitative strand
Teacher outcomes	Classroom management strategies (positive and negative)	Cycle of learning
	Collaboration with parents	Knowledge (of children's development and of classroom management theory/practice)
		Locus of control
		Emotional well-being
		Teaching philosophy and practice
Classroom outcomes	Classroom environment	Relationships (teacher-child and child-child)
		Classroom environment
Child outcomes	Conduct problems/behaviour difficulties	Behaviour
	Prosocial skills	Social skills
	Emotional problems	Emotional well-being
	Academic readiness	Academics
	Attendance	Outside the classroom (difference in behaviour while on playground or at home)
Parent outcomes	Attitude towards school	

*Note: Themes are listed under the strand in which they were measured or emerged. Those in **bold** received coverage across both strands.*

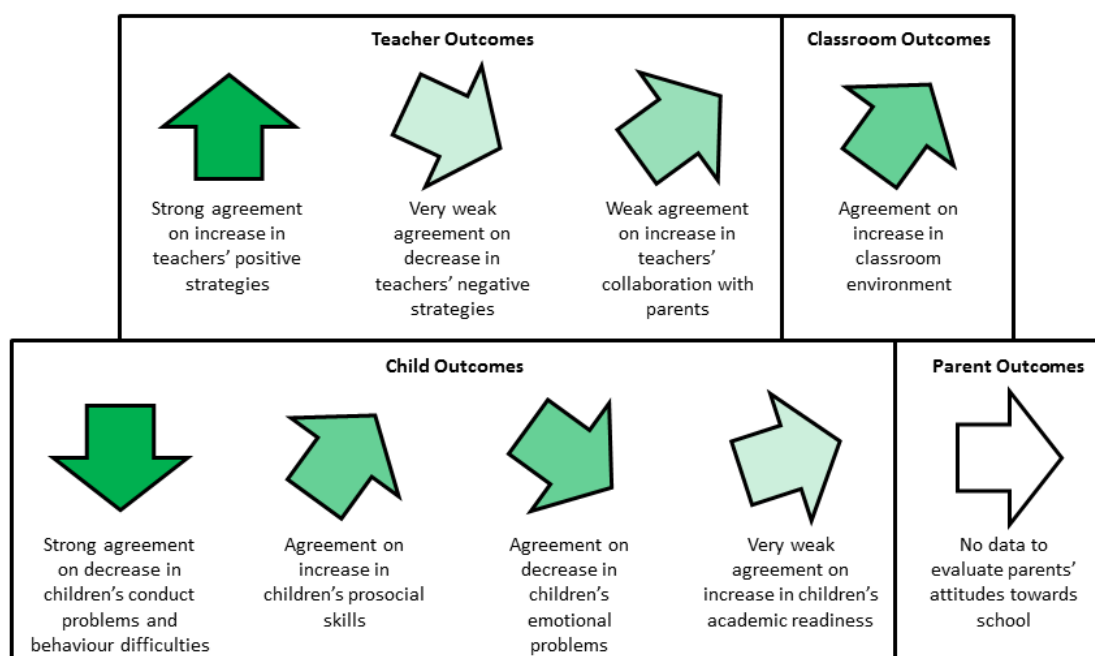


Figure 5.2 Summary of qualitative findings mapped onto quantitative constructs

Quantitative constructs	Teacher positive strategies	Teacher negative strategies	Teacher collaboration with parents	Classroom environment	Child conduct problem/behaviour difficulties	Child prosocial skills	Child emotional problems	Child academic readiness	Child attendance	Parent attitude towards school
Devon, England	Teachers learned positive strategies, including: modeling behaviour, praise, coaching, labeling behaviours and feelings, ignoring negative behaviour, greeting children and parents and using rewards, amongst others'. (Marlow et al., 2015, p. 241)				Teachers and head teachers noticed positive changes in children's behaviours					
Greater Dublin Area, Ireland	Teachers found positive strategies to be useful, including: 'praise & encouragement, use of incentives and redirections, ignoring and loss of privileges, consequences...' (Davenport & Tansey, 2009, p. 12)				Teachers noticed improved behaviour in children specifically targeted for intervention					
Limerick, Ireland	Teachers felt they had 'license' to ignore minor misbehaviour (McGilloway et al., 2010, p. 18)	Teachers expressed a decreased reliance on corporal punishment and other negative strategies such as shouting and name-calling' (Baker-Henningham & Walker, 2009, p. 634)	Teachers felt that 'they were making more effort to contact and share information with parents about their child's progress.' (McGilloway et al., 2010, p. 18)	The general atmosphere within the classroom was considered to be calmer and more co-operative, whilst it was also regarded as a happier and more positive environment in which to work (and learn)'. (McGilloway et al., 2010, p. 19)	Child behaviour improved across a range of areas, including: 'children were displaying a greater ability to stay on task, were less distracted, and were engaging in less attention-seeking behaviours.' (McGilloway et al., 2010, p. 18)	Children worked better with their peers and did not need to have adults intervene to resolve conflicts as they were able to do this themselves Children had developed better peer relations, they had become more prosocial, and showed a greater willingness to help and encourage each other.' (McGilloway et al., 2010, p. 19)	Children were better able to identify how they were feeling and express their emotions			
Kingston, Jamaica (1)	Teachers used more positive strategies, such as praising, incentives, ignoring, time-out, and others	Teachers expressed a decreased reliance on corporal punishment and other negative strategies such as shouting and name-calling' (Baker-Henningham & Walker, 2009, p. 634)	Teachers felt that in focusing on the positive rather than the negative, their relationship with parents improved	Classrooms were quieter and had less overall aggression	Child behaviour improved among both children who demonstrated problematic behaviours at the beginning of the year and those who were shy	Children shared more with their peers, helped each other more, and complimented one another	Children developed better emotional skills, including: 'better understanding of their emotions, showing more empathy, better anger management skills and showing increased confidence.' (Baker-Henningham & Walker, 2009, p. 635)	Teachers reported that children participated in lessons more, enjoyed school more, and improved their academic skills	Teachers reported that children's attendance improved	
North Wales (1)	Teachers reported learning new positive strategies and used existing positive strategies more in the classroom Positive strategies used by teachers include: praise, 'give me five' (to get children to sit quietly), and reward charts		Teachers felt that the class as a whole benefited	Teachers felt that the class as a whole benefited	Children with behaviour difficulties were more manageable to teach Children improved in their ability to pay attention	Children were more considerate of each other	Children were more confident			

Figure 5.3 Integrative grid – quantitative constructs and qualitative studies

Teacher outcomes. All five qualitative studies reported findings related to teachers' improved skills in positive classroom management strategies. These qualitative reports converge with the results from the multilevel meta-analysis of a moderate effect of IY TCM on increasing positive strategies ($g = 0.44$). In contrast, only one qualitative study (Kingston (1)) reported teachers' experiences reflecting on their decreased use of negative classroom management strategies, even though quantitative results showed a small effect of IY TCM on decreasing negative strategies ($g = -0.32$). These data raise questions about why there is such a difference in attention to positive/negative strategies in the qualitative data. At the most basic, individual level, it is possible that teachers were more aware of the presence of positive strategies than the absence of negative strategies in their day-to-day practice, which in turn affected their accounts of any benefits from the course. Alternatively, this awareness could have resulted from the phrasing of the questions or how researchers elicited responses on topics. Indeed, it is also possible that researchers might have attached greater explanatory power to data on positive strategies than negative ones during the analysis.

One reason for why the Kingston (1) qualitative study might differ from the others in its focus on decreased negative strategies is the cultural context whereby corporal punishment was an acceptable and commonly used teacher strategy at baseline. In all of the other contexts (England, Ireland, and Wales), corporal punishment in the school setting was not permitted at the time of intervention. Instead, negative strategies prevalent in these studies would have been limited to shouting, exclusion from the classroom, using indirect commands, and not allowing children sufficient time to comply with a directive before repeating it. Interestingly, even though the Kingston (1) qualitative study reports both teachers' experiences of positive and negative strategies changing as a result of intervention, the Kingston (2) quantitative study did not report on any teacher outcomes.

If this qualitative evidence is any indication of what could be measured quantitatively, the effect size for decreasing negative strategies would only be larger were such data available from the Kingston (2) study.

Teacher collaboration with parents, which was measured only by one quantitative study (Rhode Island) was addressed in the qualitative data from two studies (Limerick and Kingston (1)). The quantitative study found no significant effect in the change of teacher collaboration with parents. Teachers in the two qualitative studies felt that their relationships with parents had improved, by communicating with parents more frequently and focusing on positive aspects of their children. It is unclear whether parents also felt that their relationship with their child's teacher had improved, as parents were not interviewed or otherwise surveyed on this outcome. The scarcity of data on this outcome is a noticeable gap in the literature given that the programme is described as focusing on 'ways teachers can effectively collaborate with and support parents' school involvement, and promote home-to-school consistency' (The Incredible Years, 2013).

Classroom outcomes. Three qualitative studies (Limerick, Kingston (1), and North Wales (1)) examined teacher reports of classroom environment, and what little data there are from these studies suggests that teachers experienced improved atmosphere in the classroom, with less noise and aggression and more cooperation as a result of intervention. This is in line with trends from the two quantitative studies (Rhode Island and North Carolina) that measured classroom environment. The Rhode Island study concluded that the 'difference between [IY TCM] and comparison classrooms in amount of change [in classroom environment] was significant, and a large effect size was achieved (Cohen's $d=2.65$)' (Shepard et al., 2008). The North Carolina study found that positive climate was statistically improved but negative climate did not reach statistical significance. Here,

again, the data does not contradict itself although any conclusions should be couched in the relative dearth of evidence available.

Child outcomes. All five qualitative studies reported on teachers' experiences of benefits for children's behaviour, converging with the conduct problems and behaviour difficulties constructs from the multilevel meta-analysis. Details in the qualitative reports were not generally detailed enough to distinguish between conduct problems (i.e., aggressive/noncompliant) and behaviour difficulties more widely (i.e., including inattentiveness, peer problems), which is why these two constructs from the quantitative strand were collapsed into one column for the cross-synthesis. Generally, teachers' reports referred to 'behaviour' improving, although specific reports of improved attention among children were evident, suggesting that teachers regard attention as a valuable component of child behaviour. Given that of the two meta-analyses, the broader behaviour difficulties analysis was conducted as a *post hoc* model, this qualitative data lends credence to approaching child behaviour from this wider perspective (i.e., including attention problems) as being more in tune with how teachers assess the effectiveness of the programme.

Furthermore, teachers felt that implementing IY TCM strategies was beneficial for children who started with higher levels of problematic behaviour, supporting the findings from the multilevel meta-analysis that effect sizes for this high risk subgroup were about three times greater than for children across the spectrum of behaviour needs ($g_{all} = -0.11$ and -0.12 versus $g_{high\ risk} = -0.33$ and -0.37).

With regards to child prosocial skills, the three studies with the most breadth (Limerick, Kingston (1), and North Wales (1)) all addressed teachers' views on any benefits of IY TCM, concluding that children were more prosocial. In contrast with teachers' comments on child behaviour, there was no explicit link drawn between children

who were least prosocial at the start of the year benefiting most from the intervention. Instead, the findings were described more inclusively, such that children in general benefited prosocially. These findings are compatible with the results of the meta-analysis, which found improvements in prosocial skills ($g = 0.13 - 0.31$).

Other child outcomes measured in the quantitative studies but which were not included in any multilevel meta-analysis model include decreasing children's emotional problems and increasing their academic readiness and attendance. The qualitative evidence (Limerick, Kingston (1), and North Wales (1)) on children's emotional problems reveals improvements in identifying and expressing emotions and growing in self-confidence. The very limited quantitative evidence (Limerick, North Carolina) on emotional problems suggests small effect sizes ($g = -0.16$ to -0.30), trending, although not reaching, statistical significance. Given the limited quantitative data to power an analysis on this outcome, it is unsurprising that this outcome was not statistically significant. However, its movement in the desired direction indicates preliminary evidence of benefits due to intervention.

Even more underdeveloped are the quantitative constructs of children's academic readiness and attendance. Quantitatively, child academic readiness was only measured in one study (North Carolina) and the available findings from this study were preliminary ones presented in a conference. As such, conclusions should not be drawn until the final report is complete and analyses all of the relevant data. Qualitatively, only one study (Kingston (1)) reported findings on this construct, with teachers generally expressing their observations of children increasing their enjoyment in lessons and also improving in academic skills. Child attendance was quantitatively measured in one study (Kingston (2)) and qualitatively measured in one study (Kingston (1)). While teachers reported that children's attendance improved, the high medians and large ranges (90 [48-100]

intervention; 88 [30-100] control) make any quantitative conclusions difficult. Most studies did not address either child academic readiness or attendance in their quantitative or qualitative data. Since academic readiness was a secondary outcome in the quantitative strand, it is possible that researchers viewed the behavioural, social, and emotional outcomes as the main, immediate outcomes from IY TCM intervention, with any desired academic outcomes as more distal. Similarly, in the qualitative data, teachers might not have attributed a direct or strong enough causal link between any observed academic growth and the intervention, resulting in its omission from their experiences.

Overall, the cross-synthesis illustrates that for child outcomes, areas with the most quantitative and qualitative evidence (behaviour and prosocial) produce convergent findings, which serve to strengthen both the conclusions we can draw from these findings as well as speak to the usefulness of these conclusions for decision-makers. Areas where there is less evidence across either strand (i.e., emotional, academic, attendance) either need to receive more attention if they are considered valuable domains related to this intervention, or perhaps resources should not be directed to collecting data on these outcomes if they are only tangential and not central to the intervention.

Parent outcomes. One quantitative study (Kingston (1)) collected data from parents at post-intervention on their attitudes towards school. These findings were reported as median (range) and are therefore difficult to interpret other than to say that the similar medians and wide ranges (28.0 [5-39] intervention; 27.5 [12-37] control) overlap. Notably, none of the qualitative studies included data from parents' experiences or views of the IY TCM programme, signifying a complete lack of inclusion of that key (albeit indirect) stakeholder group in the research literature. Particularly given the earlier discussion on teacher-parent collaboration as a desired outcome of the programme, the absence of parent input – quantitatively or qualitatively – illustrates a major gap in our

154

understanding of the potential benefits of this intervention. Whether future research chooses to collect data on parents' attitudes towards school is up for debate, as it would appear that the more central parent-related outcome should be teacher-parent collaboration, since this is what is encouraged through IY TCM. However, qualitative evaluations moving forward also should seek out participation from parents across these identified constructs.

Mapping the quantitative findings onto the qualitative constructs. When the quantitative findings were mapped onto the qualitative constructs, the most apparent trend is that the data either overlap across all studies or are absent across all studies (see Figures 5.4 and 5.5). Notably, the qualitative constructs cover a much broader range of themes than is present among the quantitative findings. Of particular note, the quantitative data does not cover possible mechanisms of change or the learning process experienced by teachers participating on the IY TCM training. As a result, the many themes discussed in the cycle of learning meta-synthesis are collapsed into just one overarching category for this cross-synthesis. Similarly, only one area of teacher outcomes identified in the qualitative strand was addressed through the quantitative findings. The findings are reported in five categories: 1) cycle of learning, 2) classroom outcomes, 3) teacher outcomes, 4) child outcomes, and 5) critiques (see Figures 5.4 and 5.5 for integrative grids and Figures 5.6 and 5.7 for summaries).

Qualitative constructs	Cycle of learning	Teacher knowledge	Teacher locus of control	Teacher emotional well-being	Teacher philosophy and practice	Classroom relationships	Classroom environment	Child problematic behaviour	Child social skills	Child emotional well-being	Child academics	Child in other environments
Multi-level meta-analysis					Teacher negative strategies: $g = -0.32$ (-0.62, -0.01) Teacher positive strategies: $g = 0.44$ (0.13, 0.74)			Child conduct problems: $g = -0.20$ (-0.36, -0.07) Child conduct problems (heterogeneity): $g = -0.11$ (-0.32, 0.00) Child conduct problems (high heterogeneity): $g = -0.33$ (-0.47, -0.18) Child behaviour difficulties: $g = -0.20$ (-0.35, -0.04) Child behaviour difficulties (heterogeneity): $g = -0.12$ (-0.23, 0.01) Child behaviour difficulties (high heterogeneity): $g = -0.37$ (-0.46, -0.28)	Child prosocial skills (high heterogeneity): $g = 0.31$ (-0.03, 0.65) Child prosocial skills (heterogeneity): $g = 0.13$ (0.02, 0.24)	Child emotional symptoms: $g = -0.17$ (-0.45, 0.00) Child emotional symptoms (high heterogeneity): $g = -0.15$ (-0.44, 0.00)		
					see multilevel meta-analysis			see multilevel meta-analysis	see multilevel meta-analysis			
Limerick, Ireland					see multilevel meta-analysis			see multilevel meta-analysis	see multilevel meta-analysis			
Kingsley, Jamaica (2)					see multilevel meta-analysis			see multilevel meta-analysis	see multilevel meta-analysis			
Rhode Island, USA					No significant difference between IV TCM and comparison group on change in use of inappropriate strategies Although the amount of change in conflict scores did not differ in the two groups, $F(1, 102) = 16.76$, $p = 0.07$ (IV TCM classrooms maintained improvement) (Shepard et al., 2008) The difference between IV TCM and comparison classrooms in amount of change was significant, and a large effect size was achieved (Cohen's $d = 2.65$) (Shepard et al., 2008)			There was no group difference in either problem behaviour scores at post-test or on changes in problem behaviour over the course of the year ($F(1, 107) = 2.71$, $p = 0.11$). Among children with an elevated score at baseline (> 85), 50% of children in IV TCM classrooms and 75% of children in comparison classrooms demonstrated greater reduction in their score. (Shepard et al., 2008)	Little improvement was demonstrated on children's prosocial or cooperative behavior, or on peer rejection, regardless of the year ($F(1, 107) = 2.71$, $p = 0.11$). Among children with an elevated score at baseline (> 85), 50% of children in IV TCM classrooms and 75% of children in comparison classrooms demonstrated greater reduction in their score. (Shepard et al., 2008)			
North Carolina, USA					see multilevel meta-analysis			see multilevel meta-analysis	see multilevel meta-analysis			
North Wales (2)					see multilevel meta-analysis			see multilevel meta-analysis	see multilevel meta-analysis			
					see multilevel meta-analysis			see multilevel meta-analysis	see multilevel meta-analysis			

Figure 5.4 Integrative grid – qualitative constructs and quantitative studies, part one

Qualitative constructs (subthemes)	Motivation for taking course	IV TCM process and content			Child outcomes			Wider support and scaling-up	
	Compulsory vs. choice	Scripted vs. flexible	Culturally adapted video vignettes; omit role play etc.	Misunderstandings and disagreements	Age	SEN status	Class size	Investment from school leadership	Whole school approach vs. select teachers
Multilevel meta-analysis							Mean size: 20 children; Range: 11-20 children		Two participating teachers from each school (one in intervention, other in control)
Limerick, Ireland						Children were excluded from the study if they had a formal diagnosis of a developmental disorder [...] (McGilloway et al., 2010, p. 9)			
Kingston, Jamaica (2)	Inclusion criteria for pre-schools were: ' [...] and all teachers consented to the trial. In total, 50 pre-schools were approached and 24 meeting the inclusion criteria were recruited.' (Baker-Henningham et al., 2012, p. 101)		Supplementary video vignettes showing Jamaican classrooms were used.' (Baker-Henningham et al., 2012, p. 103)				Minimum of 20 children in each class as per inclusion criteria		24 pre-schools participated, each with three to four classrooms with a minimum of 20 children
			Additional role-plays; practical activities and small group exercises were designed based on the experiences and concerns of Jamaican teachers.' (Baker-Henningham et al., 2012, p. 103)						
			Handouts were revised to include examples relevant to the Jamaican context.' (Baker-Henningham et al., 2012, p. 103)						
Rhodes Island, USA									13 pairs of teachers and teaching assistants participated from Head Start schools
North Carolina, USA					Children in kindergarten through second grade: 'We hypothesized that students in younger grade levels [...] would experience the greatest benefit.' (Murray et al., 2014, p. 1)		Mean size: 19 children; Range: 10-24 children		Grade levels within schools were randomised to intervention or control
		Grade-level interactions tested for child social-emotional, attentional, and academic outcomes.							
North Wales (2)					Children between ages 3- and 7-years-old		Mean size: 18 children; Range: 14-30 pupils		12 classrooms from 11 primary schools participated

Figure 5.5 Integrative grid – qualitative constructs and quantitative studies, part two

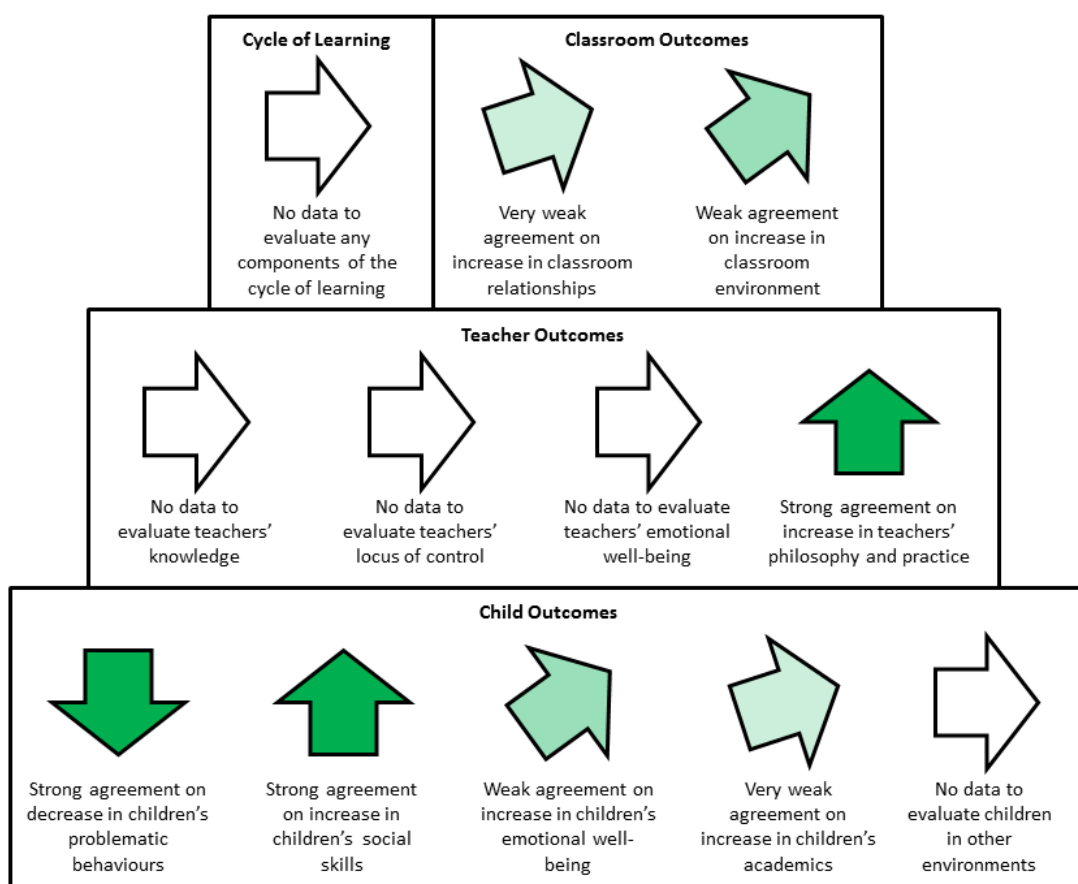


Figure 5.6 Summary of quantitative findings mapped onto qualitative constructs

Cycle of learning. In the cycle of learning findings from the qualitative meta-synthesis, it was revealed that teachers experience a repeated process of primary learning (where they learn from practice and self-reflection) and secondary learning (where they learn from others' experiences and group discussions). Throughout the journey of being on the IY TCM course, teachers felt that they were able to connect with both group facilitators and other teachers as they built up a bank of resources in classroom management, so that they were better prepared for current and future classroom scenarios. None of the quantitative studies examined this process. It is arguably more challenging to quantify the processes of an intervention compared to its outcomes of interest, and considerations in such a process include whether the components of this intervention

should be defined by their form or function. Certainly concerns about not wanting to overburden participants with data collection are real. However, the result of not collecting such quantitative data is the risk of a black box intervention. As it stands, although the outcome data is promising, we cannot conclude which components of the programme contributed most to the outcomes, or whether the primary and secondary learning were equally important, among other relevant process questions. This lack of understanding of how the intervention actually works compared to how it is theorised to work is problematic when, after a handful of positive outcome evaluations, the programme has the potential to be scaled up or transferred to new contexts. Rarely is a programme delivered to scale with the same fidelity with which it was evaluated, because adaptations are made intentionally or unintentionally. However, through understanding the core components of this intervention and being able to communicate the processes by which it has been shown to work, future decision-makers and stakeholders can make informed choices on whether the programme fits their problem and any specific contextual features they might need to adapt. The lack of quantitative data on the cycle of learning is mirrored in the teacher outcomes, as described below.

Classroom outcomes. Findings from the qualitative strand indicated that teachers noticed more positive teacher-child and child-child relationships as well as an improved and happier classroom environment. Quantitatively, these issues are weakly covered, with one study (Rhode Island) finding a trend towards reduced classroom conflict and only two studies (Rhode Island, North Carolina) calculating effect sizes that trend towards significance on better classroom environment in intervention rooms. Neither area had enough coverage or strong enough evidence to draw substantial conclusions regarding the effectiveness of the IY TCM programme in improving relationships and environment at the classroom level. However, given that there is agreement between the qualitative

evidence and these preliminary quantitative results, this appears to be an area with potential positive effects and worth consideration in future studies.

Teacher outcomes. The qualitative strand revealed four key areas in which teachers felt they had benefitted from the IY TCM programme: 1) knowledge, 2) locus of control, 3) emotional well-being, and 4) teaching philosophy and practice. Of these areas, the first three were not measured in any of the quantitative studies. Only the fourth area received coverage from quantitative studies, with all studies reporting on teaching practice. Thus, contrary to the dearth of evidence across those first three areas, there is a strong agreement of the qualitative and quantitative findings on the benefits of IY TCM training on teachers' practice in the classroom. However, the process traced through the various teacher outcomes in the qualitative strand highlighted the fact that the other three identified teacher outcomes were considered valuable by participants. Without any quantitative evidence available to ascertain the effectiveness of the IY TCM programme on teachers' knowledge, locus of control, or emotional well-being, we cannot draw any solid conclusions about variety of secondary effects this intervention might have. Even if a primary reason schools might choose to train their teachers in IY TCM remains changing teacher classroom management practices, knowing about these secondary outcomes could serve as additional motivation for decision-makers to invest in this type of professional development for their educators. Furthermore, the qualitative strand explains a possible theory of change for teacher practice, by which teachers first gain knowledge, then widen their locus of control, then emotionally feel calmer and less stressed, and finally embed the IY TCM approach into their teaching philosophy and practice. Although the qualitative strand cautioned that this was only one way to theorise the way teachers experienced these outcomes, in this explanation these first three areas would serve as mediators. In

particular, the suggestion is that increased teacher knowledge of child development and

classroom management strategies mediates changes in teacher practice. Similarly, that increased teacher locus of control and emotional well-being mediates changes in teacher practice. These are hypotheses that could be tested in future studies, although the confidence that we can place in these hypotheses is lessened for not having any preliminary quantitative evidence to indicate whether these areas of teacher development quantitatively change post-IY TCM training.

Child outcomes. Five key areas emerged in the qualitative strand regarding benefits of IY TCM for children: 1) behaviour, 2) social skills, 3) emotional well-being, 4) academics, and 5) other environments. The first three of these areas were most robustly supported in the qualitative data, with the fourth and fifth areas much less present in the available data. Similar coverage trends are seen among the quantitative studies. There is strong agreement within the quantitative studies to support the qualitative constructs of improved child behaviour and social skills. All of the studies examined these two constructs in some way, indicating the centrality of child behaviour and social skills as desired and valued outcomes from both the teachers' perspectives as well as the researchers' agendas. Children's emotional well-being was less present across the studies, with only two studies (Limerick, North Carolina) measuring it and no significant findings emerging from the data. If this construct is considered important to the classroom teachers, as the findings from the qualitative meta-synthesis suggest it is, then it would be advisable for future studies to quantitatively include measures of children's emotional well-being to contribute much needed data in this area.

There is even less evidence on the remaining two areas, with one study (North Carolina) reporting preliminary results of the effectiveness of IY TCM training on children's academic performance and no studies examining children's outcomes in other environments. The gap in reports on the effectiveness of IY TCM training on academic

outcomes is notable, considering this was an *a priori* secondary outcome identified for the quantitative strand, based on the programme’s intended purposes. The North Carolina conference abstract which reported findings on reading and mathematics achievement is too preliminary to use for decision-making (D. Murray, personal communication, 1 May 2014). However, this study’s final report is in the process of being published, and the complete academic results will serve as the first indicators of whether training teachers in classroom management strategies can affect change for children’s educational attainment. Academic attainment data will also be forthcoming from the Devon study.

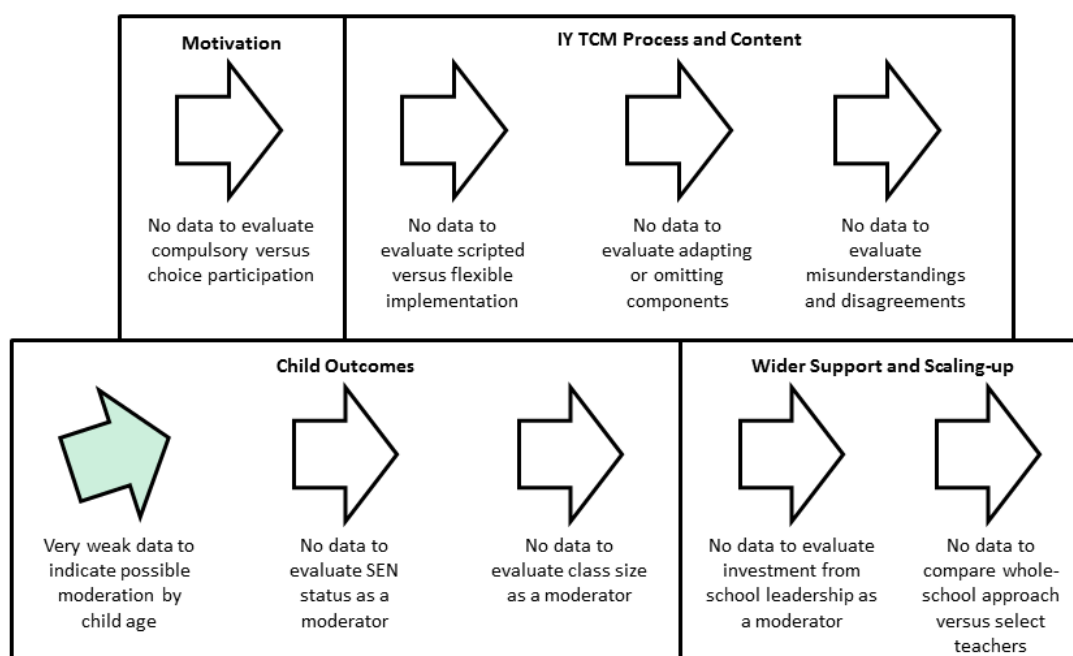


Figure 5.7 Summary of quantitative findings mapped onto qualitative construct of ‘critiques’

Critiques. Through the qualitative strand findings, several critiques of the IY TCM programme were raised by participants (mainly teachers, but also group facilitators). Although these critiques were less consistent overall across participants and between studies, they introduced important issues for discussion.

Motivation. One suggestion from the qualitative strand is the potential for participants’ motivation to act as a moderator for teacher outcomes. Specifically, the

hypothesis proposed through the critiques was that teachers who volunteered to participate in IY TCM training would experience greater benefits from the course than teachers who were mandated to participate. This hypothesis raises very interesting and relevant questions related to how we can anticipate wider scaling-up of this programme to affect outcomes. None of the quantitative studies reported controlling teacher or child outcomes based on this baseline characteristic. While the Kingston (2) study specified that all of the teachers in the included preschools consented to the trial, even this does not give enough detail to know whether some of those teachers were pressured or mandated to consent by nature of the fact that all of their colleagues were consenting and their head teacher had signed off on the project.

IY TCM process and content. The critiques around the process and content of the IY TCM programme are all situated within the fidelity-adaptation debate. Specifically, how flexible can the group facilitators be when delivering the manualised training? Interestingly, it was the facilitators who appreciated the highly structured course and the participating teachers who felt more flexibility was needed. Also, are there issues of acceptability and appropriateness that need to be addressed when transferring the programme materials to new countries? Although it is standard practice to translate the language (e.g., English to Welsh), the creation of new video vignettes was highly regarded as worthwhile, because the standard vignettes were viewed as being at odds with non-American educational contexts. However, only one of the quantitative studies (Kingston (2)) culturally adapted the video vignettes. This adaptation occurred after the initial piloting work (Kingston (1)) in which the qualitative data indicated that teachers wanted to see Jamaican classrooms portrayed in the IY TCM materials. It is unclear whether adapting the video vignettes resulted in better intervention effect sizes as this was not tested. However, it is one example of improving the acceptability of an intervention by

taking into account stakeholders' feedback. With preliminary evidence for effectiveness in a number of countries, the cultural adaptation of video vignettes would serve as a valuable next step in the process of potentially scaling up IY TCM implementation in these countries.

Child outcomes. A number of child characteristics were identified in the qualitative strand as perceived moderators of IY TCM intervention effects. Notably, themes emerged around younger children and those in smaller classes responding better to IY TCM strategies than older children and those in larger classes. Also, children identified with special educational needs (SEN) were reported as presenting additional difficulties for teachers and perhaps responding poorer to IY TCM strategies than their peers without identified SEN. With the exception of the North Carolina study, which conducted preliminary grade-level interactions for children's social, emotional, and academic outcomes, none of these possible child-level moderators were addressed in the evidence.

Wider support and scaling-up. Two related themes that emerged in the qualitative strand related to the need for school leadership to be invested in IY TCM and for the whole school to be trained (rather than just one teacher). Neither of these constructs was addressed in the quantitative analyses. No study measured the extent to which school leadership was invested in and supported IY TCM training as a potential moderator (hypothesised to increase effectiveness). Similarly, no study implemented IY TCM across an entire school, comparing whole school implementation to whole school control.

Given that the qualitative and quantitative data included in the current review generally came from research teams that conducted mixed methods evaluations at one time, it is unsurprising that the coverage of these issues within the quantitative studies was almost non-existent. To have addressed these topics within the results of the quantitative studies, the researchers would have needed to take these critiques into consideration during

study design (i.e., they would have needed to conduct a second study following the one in which these critiques were first elicited). Whether all of these critiques require extensive follow-up research remains to be seen, although following through in exploring the role of some of these topics with regards both to the effectiveness and people's experiences with IY TCM does seem the logical next step in scaling-up a service that has thus far produced desired effects and been well-received by participants.

Conclusion

This cross-synthesis drew together effectiveness and experiential results from RCT data and interview/focus group data to produce a comprehensive review of IY TCM processes and outcomes. The results of this mixed methods systematic review indicate that IY TCM is effective for both child and teacher outcomes and that the programme is acceptable and well-received by participants. In contrast to universal, school-wide programmes such as Positive Behavioural Interventions and Supports (PBIS), IY TCM can be implemented at the teacher-level; however, qualitative data in this review suggests that there is potential benefit of a systems implementation approach for IY TCM, much like what is seen among PBIS studies (Ross, Romer, & Horner, 2011; Sugai & Horner, 2006). Testing the effectiveness of IY TCM at the school-level rather than the teacher-level would incur greater costs due to the unit of randomisation (school) requiring more teachers and children included in the overall study to statistically power analyses. However, if there is theoretical justification that IY TCM would be more effective if applied within a systems implementation perspective, then this could be a next step in research. Stepped-wedge designs where IY TCM is rolled out in a staged process to schools over multiple years offer future researchers the opportunity to experimentally test the effectiveness of whole-school implementation as well as achieve longer-term outcomes. Of course, the decision

to implement IY TCM at the school-level brings the discussion back to the issue of teacher choice and motivation to participate.

Past research in this area acknowledges that although adult education is distinctive from child/adolescent education in part by the voluntary-compulsory divide, this dichotomy is far from clear-cut (Rockhill, 1983). Even if teachers were to voluntarily participate in the IY TCM programme instead of being told to do so by their head teacher, they likely have direct (e.g., employer rules or governmental licensure laws) or indirect (e.g., union rules) regulations to which they must adhere (Rockhill, 1983). These regulations could serve as one form of compulsion, even if they exist with the intention to protect children's learning experiences by ensuring that educators are knowledgeable about new developments in their field. Thus, the existence of universal obligation to participate in training courses such as the IY TCM programme could negate the more localised pressures (e.g., orders from head teacher) facing only some of the participating teachers. A moderator effect regarding motivation for taking the course would have implications for anticipating the effectiveness of scaling-up implementation of this programme to whole schools or regions, with appropriate reductions in predicted outcomes.

Related to how IY TCM might be scaled up to teachers and schools more broadly, concerns over (un)intentional adaptations need to be addressed. Those who are hesitant to adapt interventions cite apprehensions stemming from our lack of understanding about core components and the subsequent fear of omitting or changing core components (Elliott & Mihalic, 2004; Mihalic, 2002). However, the opposite argument in favour of cultural adaptations recognises that adaptations can increase intervention acceptability and increase the effectiveness of the intervention among new groups (Lau, 2006). Certainly, any adaptations of the video vignettes to a new country would require an influx of resources,

including teachers/children who would be filmed in various scenarios as well as the finances and time to successfully carry out the filming. Yet, were these adaptations to be carried out only in form – like the language translations – and not include changing the function (i.e., intended messages) of the vignettes or the discussion points associated with each clip, the resulting hypothesis is that participants would perceive the vignettes to be more acceptable and relevant to their own classroom practice and that any change in effectiveness would favour the intervention group more than current evidence. In other words, it is possible that the meta-analysis effect sizes are underestimated, particularly given that more than half of the included studies were conducted outside of the country in which the video vignettes were made. With the available effectiveness data and encouraging qualitative experiential data, a convincing argument for investing the necessary resources for developing new video vignettes could be made prior to broader implementation of the programme.

Additional resource constraints are relevant to how and for whom IY TCM works. Class size is one area that is often beyond individual teachers' and schools' control. Despite evidence to demonstrate that small teacher-pupil ratios are more effective for children's development in school, particularly for children identified with SEN or those who are from ethnic minority backgrounds (Schanzenbach, 2014), reducing class size is no simple feat. Even when class size is reduced, teachers are not necessarily prepared to change their practices in such a way that affects children's outcomes (Graue, Hatch, Rao, & Oen, 2007). Rather, professional development opportunities are still required to support teachers in these smaller classroom settings (Graue et al., 2007). As a result, the case could be made that whether teachers have large (e.g., 30) or small (e.g., 15) class sizes, both could benefit from professional development to support their effectiveness. With regards to the IY TCM programme, a relevant question that emerges is whether delivery of

the course could be better tailored to teachers' needs were it stratified by class size. Specifically, should IY TCM be delivered in groups where participating teachers either all have large class sizes or all have small class sizes (rather than without regard to class size) in an effort to allow for bespoke discussion around techniques used that will maximise effectiveness given the different ratios of teachers to pupils?

Finally, while there is compelling evidence that expanding IY TCM training would be beneficial to teachers and children alike, not all children benefited and certain children were identified as potentially distinct subgroups. As previously discussed in Chapter 1, there has been a global shift towards SEN inclusion in the past decades, particularly following the Salamanca Statement (UNESCO, 1994). Despite this international agreement, much work remains to be done in preparing and supporting teachers in inclusion settings. Teachers must have both a positive mind-set that favours inclusion as well as the skills required to 'make it work' (Winter, 2006, p. 85). Even though the landscape of classrooms has been changing for decades, there does not appear to have been similarly paced change in the initial teacher training programmes to alleviate teachers' concerns over supporting these children identified with SEN in mainstream settings (Winter, 2006). Among the included quantitative studies in this review, none addressed the possible moderating effect of children's SEN status on outcomes. Furthermore, the Limerick and Kingston studies explicitly excluded children with formal diagnoses of developmental disorders despite the fact that the need to exclude these children indicated their presence in the included classrooms. Thus, although it was likely that learning strategies that could also apply to children with SEN would have been desirable and relevant to these teachers, we have no information on whether training these classroom teachers in IY TCM was effective for the children with SEN in their classrooms. Whereas other child characteristics not raised in the qualitative data – such as qualifying for

free/reduced school meals (as a proxy of family socio-economic status) and native language (if not the same as instruction) – were used in the North Carolina study to explore possible interactions, SEN, which was raised as a concern, remains unstudied in relation to the IY TCM programme.

This dissertation now proceeds to Study Two, which qualitatively explores this identified concern over the subgroup of children identified with SEN. Since this review has highlighted, among other findings, that children's SEN status might moderate intervention effects, Study Two asks educators responsible for SEN provision in their schools about effective classroom behaviour management strategies for children identified with SEN. Just as qualitative research was a first step when culturally adapting an intervention for a population in a new context, as done in Jamaica with IY TCM (Baker-Henningham, 2011), Study Two seeks to understand whether IY TCM would be acceptable and appropriate for a new population (children identified with SEN).

STUDY TWO

Classroom behaviour management strategies in response to problematic behaviours of
primary school children with special educational needs: views of special educational needs
coordinators

Chapter 6 : Methods

Need for the Current Study

Schools serve children with a variety of different needs. They might have children who display problematic behaviours in the classroom (e.g., conduct disorders), and schools might have children with identified special educational needs (SEN). Teachers and parents in Great Britain identify around 17% of children at age seven as having SEN (Parsons & Platt, 2013), and a national survey of mental health of children and young people in Great Britain estimate the prevalence of conduct disorder (CD) at roughly 6% of school-aged children (H. Green et al., 2005). Almost 60% of children with CD are behind in school (with over a third of children with CD lagging two or more academic years), and it is estimated that 52% of children with CD are also identified by their teachers as having SEN (see Figure 6.1) (H. Green et al., 2005). It is likely that schools will face additional challenges supporting children with such complex needs.

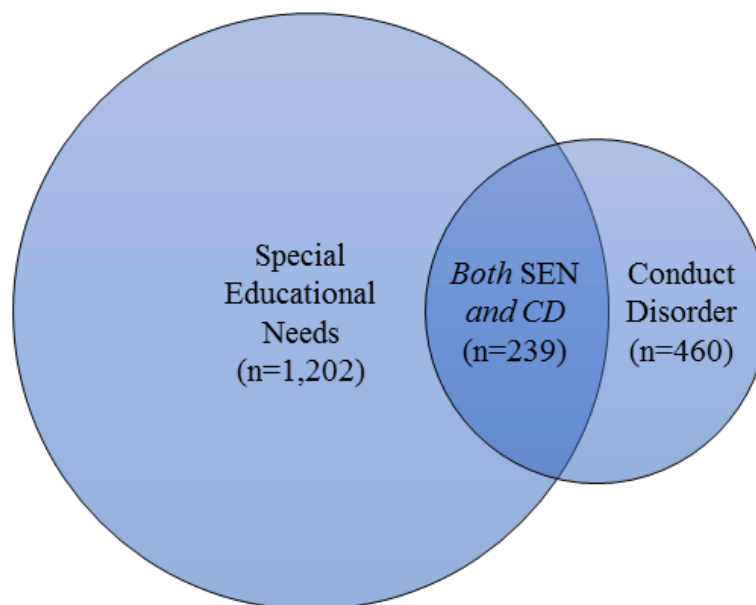


Figure 6.1 Subgroup of interest for questions in qualitative interviews: children identified with both SEN (Parsons & Platt, 2013) and CD (H. Green et al., 2005)

The IY TCM programme is especially relevant to the behaviour management of children with special behavioural needs, as demonstrated in Study One. Although children with special behavioural needs have been studied with regards to IY TCM, studies have not examined whether these effects hold for the subset of children with special behavioural needs who are also identified with SEN. And yet, Study One also highlighted the qualitative research suggesting that teachers experience additional difficulty when implementing IY TCM management strategies when including children with SEN in mainstream classrooms. Indeed, since schools will have access to different internal and external resources to provide extra support to their children with SEN (e.g., one-to-one assistants and emotional curricula), it is important to know more about how schools approach behaviour management with this particular subgroup of children.

The questions in these interviews are intended to examine how schools work with children who have *both* special educational needs *and* special behavioural problems. In this way, it serves a preliminary albeit critical role in the iterative research process of understanding for whom IY TCM might be appropriate and whether there are any important considerations needed (such as adaptations) to implement IY TCM with this population. One of the primary benefits of qualitative interviews is the opportunity for participants to address issues that are important from their perspectives, rather than feeling constrained by issues pre-determined by the researcher (Barbour, 2014). Particularly, semi-structured interviews provide for increased flexibility during the interview process. Following the current advice from the Research Excellence Framework in the UK, researchers should extend their work beyond immediate clients and service users to include other stakeholders such as practitioners and service planners (Barbour, 2014).

Qualitative research with teachers, parents, and other stakeholders has been conducted on previous IY TCM trials in other contexts, as previously meta-synthesised in

Chapter 4 (Baker-Henningham & Walker, 2009; Davenport & Tansey, 2009; Hutchings et al., 2007; Hyland, 2014; Marlow et al., 2015; McGilloway et al., 2010; McGilloway et al., 2012). While these data relate to children with behavioural problems, most of the research does not directly discuss children with special educational needs. Responses elicited from teachers that have indirectly touched on this subgroup have highlighted the additional difficulties in implementing the TCM strategies in mainstream classes that include students with special educational needs (McGilloway et al., 2012). These data suggest that further research looking specifically at the impact of IY TCM on children with special educational needs—and doing so through the school teachers/staff with special educational assignments (e.g., SENCOs, TAs)—is currently needed to fill a gap in this body of literature.

As introduced in Chapter 1, this qualitative interview study asks questions to SENCOs as part of a process evaluation that parallels the on-going STARS Study, which is an 80-school cluster RCT of IY TCM in England (Ford et al., 2012; Hansford et al., 2015). Furthermore, these interviews will serve to better understand contextual factors, such as the range of support services offered to children with special educational needs through the schools.

Steckler, Linnan, and Israel (2002) have identified seven key components for process evaluation: context, reach, dose delivered, dose received, fidelity, implementation and recruitment, which may be of interest when evaluating trials. Recent commentary on the 2008 MRC guidance on process evaluation suggests that extending the guidance to consider theory-driven approaches, which assess whether an intervention works by focusing on the assumptions about why it is thought to work, can help illuminate contextual influences on uptake and utility as well as variation in outcomes across different groups or circumstances (Anderson, 2008; Freeman, 2009) and this sits well with a realist

approach (Greenhalgh et al., 2009; Pawson & Tilley, 1997). In developing the qualitative component of the process evaluation to parallel the large cluster randomised controlled trial, the aim was to identify experiences, contexts and processes within the on-going trial, and the ways in which these relate to change in teachers, with impact for children. In this sense, the approach aims to identify influencing mechanisms (processes) and contexts (personal, professional and organisational) in the use of IY TCM strategies (barriers and facilitators).

Aims

The aims of this study were to explore SENCo's experiences and views on supporting children identified with SEN. To this end, the research questions included:

1. What classroom behaviour management strategies do SENCo's find to be successful when working with children identified with SEN who also have behavioural difficulties?
2. Are the SENCo-generated strategies compatible with the IY TCM content such that the course could be considered acceptable and appropriate for educators who support this subgroup of children?

Study Design

The STARS Study design includes a pilot phase (Cohort 1), followed by the main phase (Cohorts 2 and 3). Each school is in the study for three years: the first year of recruitment followed by two follow-up years. Data were collected at baseline (0 month), post-intervention (9 month), first follow-up (18 month), and second follow-up (30 month). The qualitative process evaluation design parallels these phases. These qualitative interviews with SENCo's occurred within the main phase of the trial, at post-intervention with Cohorts 2 and 3.

Sample

All 15 SENCOs from Cohort 2 intervention group schools were invited to participate in the telephone interviews. Such an approach allowed for data to reflect the range within this group of SENCOs rather than selecting characteristic cases (Barbour, 2014). Due to concerns over contaminating the control group with questions exploring classroom behaviour management strategies and because some of the questions specifically targeted IY TCM knowledge, no interviews were conducted with control SENCOs until after the targeted teacher in the school received IY TCM training in the following academic year. Therefore, invitations were subsequently sent to the 15 Cohort 3 intervention group SENCOs and 10 control group SENCOs from Cohort 2, whose school's target teachers had since received the intervention. Although not the only issue affecting sample size in qualitative research, a guiding principle includes the concept of saturation (Guest, Bunce, & Johnson, 2006; Mason, 2010). With a potential of up to 40 SENCOs to interview, this study's maximum sample size took account of current guidance on optimising sample size in qualitative research.

Data Collection and Management

Data collection during the qualitative telephone interviews with SENCOs ran parallel with the main phase of the STARS trial. The semi-structured interviews respected both the aims of the qualitative study and the practicalities of data collection with busy professionals. Information on the qualitative methods was given to participants both at recruitment and with the invitation to participate in the interviews, as will be described in further detail below.

Semi-structured interviews

Data was collected with SENCoS through telephone interviews, an advisable technique when working with participants scattered across a large geographical area (Barbour, 2014). Compared to in-person interviews, telephone interviews are typically shorter in length and use a semi-structured format, even though both types of interviews may take the same form. Benefits of utilizing telephone interviews instead of in-person interviews include improved participation among busy professionals and facilitating a focus on evaluation questions, while simultaneously allowing participants to share their experiences and views (Padgett, 2008). Telephone interviews serve as useful means to gather information about people's experiences and views with minimal imposition (Sturges & Hanrahan, 2004). In this study, telephone interviews lasted a maximum of 30 minutes and were pre-planned to optimise participation and privacy.

Topic guide questions. In an effort to conduct the interview with maximum efficiency to respect the busy schedule of the SENCoS, a topic guide (Appendix F) was sent out to in advance, allowing them to look up relevant routine demographic and school information (Barbour, 2014). The prompts on the topic guide were ordered from more general questions to specific probes, as it is recommended when interviewing to gradually move towards more probing questions after starting with less direct ones (Barbour, 2014).

The topic guides covered the following areas:

- background information about the SENCo, their role and the level of SEN in their school;
- their general view of SEN and behaviour management;
- how they assist teachers of children with SEN;

- their views of the TCM intervention; and
- information about SEN services used by their school.

The researcher asked all interviewees the same questions in roughly the same order and followed a purpose-designed interview guide with a list of questions and prompts appropriate to the stage of research.

Interviews were digitally audio-recorded and transcribed for subsequent analysis. Participating SENCOs were reminded of the audio-recording and asked to provide verbal consent at the beginning of the interview according to a script. Transcribing interviews allows for more accurate future analysis of the data, but care must be taken when inserting punctuation into the transcript, as punctuation, tone, and emphasis can all affect the meaning of the words (Barbour, 2014). It is advised that when possible the researcher conducting the interview should also be the one to transcribe it and that the transcript should be read while simultaneously listening to the recording to pick up on any such additional cues for meaning (Barbour, 2014). Therefore, the researcher who conducted the interviews also transcribed and analysed them in this study.

Procedure

Beginning of academic year. All SENCOs received an information leaflet at the beginning of the academic year that their school is taking part in the study. This leaflet described the purpose of the STARS Study and outlined the ways in which they might be asked to participate. It asked for SENCOs to provide information on pupils with SEN in the target classroom and explained that the SENCOs might be contacted later in the academic year to participate in a telephone interview, which would occur during the summer term.

Spring term. All SENCOs from schools in the intervention group received an information sheet (Appendix G) inviting them to participate in a telephone interview. The information sheet again described the STARS Study, explained the purpose of the telephone interviews, and asked that they consider participating in a telephone interview. SENCOs who agreed to take part in the telephone interviews were contacted via email to arrange a suitable date/time for the interview to occur. This email also included the topic guide so that the SENCOs were able to look over the questions in advance, if desired.

Summer term. Telephone interviews with the SENCOs took place during the Summer term of the academic year during which the target teacher received TCM training. After the interview has taken place, SENCOs received a £10 gift voucher in appreciation of their participation.

Consent and ethics

These qualitative interviews had the potential to elicit sensitive information and confidential data, and as such researcher discretion and adherence to approved ethical practices was of utmost importance. Throughout the process of recruiting SENCOs for the telephone interviews and conducting the interviews, it was stressed that data would be kept confidential, stored securely, and any findings published or otherwise disseminated as a result of these interviews would remain strictly anonymous. SENCOs had the opportunity to read through the information sheet as well as ask any clarifying questions to the research team well in advance of participating in the telephone interviews. Verbal consent to participate and record the telephone interviews was requested at the start of each interview (see Appendix H for standard operating procedure). All participants were clearly informed that they had the right to withdraw their personal data from the study at any time and without consequence.

This study was approved by two university ethics committees: 1) the University of Exeter Medical School (which also approved and oversaw the larger STARS Study) and 2) the Departmental Research Ethics Committee (DREC) in the Department of Social Policy and Intervention, University of Oxford (Ref: 2013-14_56; see Appendix I).

Analysis

Analysis of the data collected from these qualitative telephone interviews was conducted from verbatim and anonymised transcripts. The same researcher who conducted the interviews also transcribed them, so as to best account for appropriate tone, emphasis, and punctuation in the transcripts and to prevent inserting bias or altering the meaning of the words (Barbour, 2014). Data was stored on a password-protected computer.

Analysis was guided by a realist perspective, to identify experiences as the lived ‘reality’ of participants. Thematic analysis of interview and focus group data was both theoretically driven by the research questions and allowed for more inductive analysis whereby emergent themes were also identified. In the analysis, ‘keyness’ of themes did not necessarily relate to incidence of occurrence but to whether a theme captured information relevant to the research questions, in this case relating to a range of trial processes (Braun & Clarke, 2006). This approach helped to explicate patterns of experience and views of SENCos.

Thematic analysis was conducted using data from the participating Cohort 2 (intervention) SENCos in the Summer Term when it was collected. The following year, data from the participating Cohort 3 (intervention) and Cohort 2 (control) SENCos was analysed deductively using the themes generated from the first round of analysis. However, this second round of analysis allowed for the development of new themes to be then considered against the round one data.

Chapter 7 : Acceptability and applicability of a teacher classroom management programme for children with special educational needs in mainstream schools –

Results from semi-structured qualitative telephone interviews

Descriptive findings

Six SENCOs from South West England participated in the first round of this study. The SENCOs were all experienced teachers with a mean average of 24 years teaching (range: 14 – 32 years). Furthermore, the participants had served as their school SENCO for a mean average of 7 years (range: 2 – 15 years). Of those SENCOs who were invited to participate but did not interview, three agreed to participate but were unreachable despite multiple attempts, one declined to participate due to time-constraints, and five never responded to researchers' invitations.

Nine SENCOs participated in the second round of this study. They had a mean average of 18 years teaching (range: 2 – 30 years). They had served as their school SENCO for a mean average of 6 years (range: 0.5 – 15 years). Of those SENCOs who were invited to participate but did not interview, one agreed to participate but was unreachable despite multiple attempts, four declined to participate due to time-constraints and lack of interest, and nine never responded to researchers' invitations. Two school SENCOs were not contacted regarding interviews due to staff changes leading to the target teacher or head teacher taking over SENCO duties and therefore already being invited to participate in either teacher focus groups or head teacher interviews.

Overall, the SENCOs across both rounds had a mean average of over 20 years teaching (range: 2 – 32 years) with a mean average of over 6 years as SENCO (range: 0.5 – 15 years). Information was not available on non-participants and so it cannot be determined how the participants and non-participants compared to each other.

Overview of findings

The subsequent sections first cover the types of problematic behaviours described by the SENCoS and then address their approach to determining the cause of problematic behaviours before exploring the SENCoS' approaches to classroom behaviour management (see Figure 7.1).

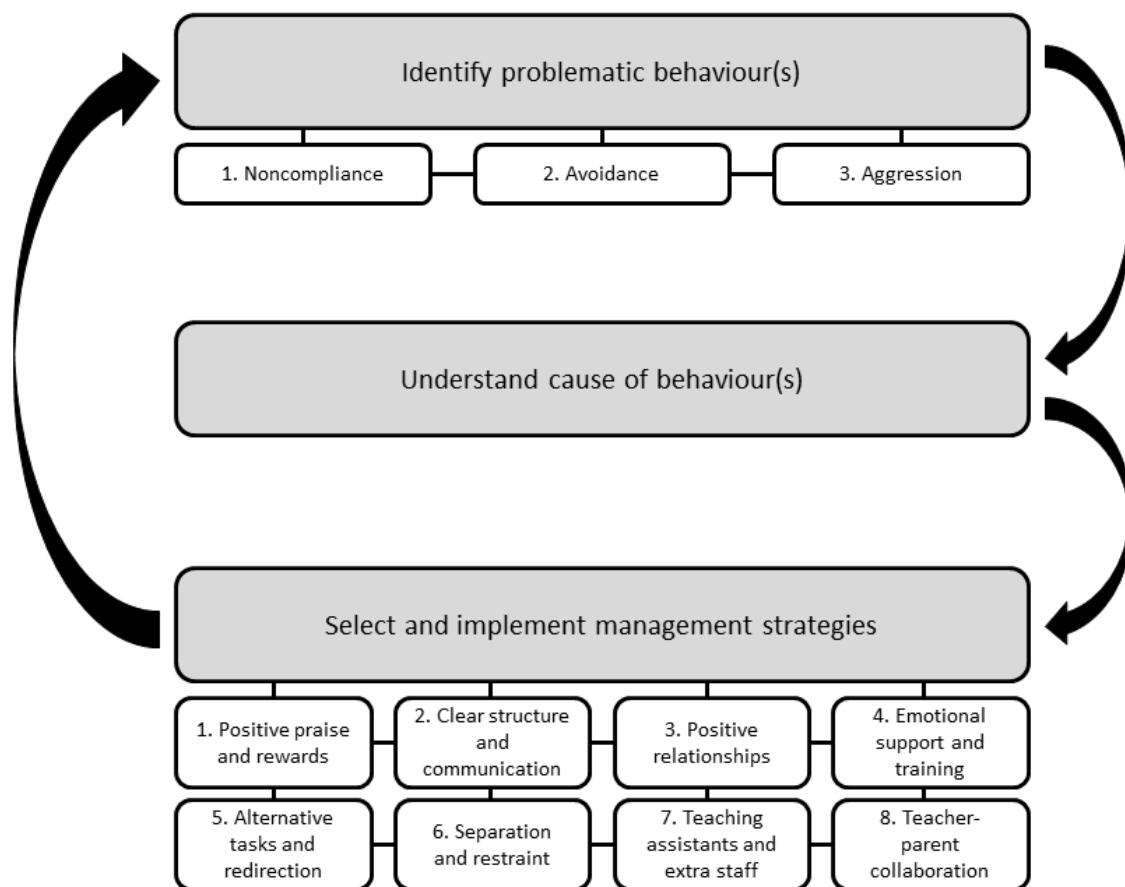


Figure 7.1 Overview of findings from SENCo interviews

Although it should be understood that teachers and school staff are implementing strategies for both prevention and treatment purposes and that this is an iterative process, because the questions in this study were framed to address problematic behaviours, the findings are presented in a manner that highlights the reactive process (starting with problematic behaviours) although many of the strategies also serve to proactively prevent

these situations in the first place. An example of the coding framework can be found in Table 7.1, showing the many types of sub-codes generated during open coding, which were then grouped into coding categories and then under themes. After detailing the strategies generated by the SENCOs interviewed in this study, the strategies are presented mapped alongside the IY TCM content to address how they compare.

Table 7.1 Example of coding tree for management strategies

Themes	Codes	Sub-codes
Management strategies	Positive praise and rewards	Use positive approach, reward systems, teacher praise for child, hook child's interest, other key adults praise child
	Clear structure and communication	Daily lesson plan written up for children to see, use of visuals, numbered to-do lists, children tick off items, children know day's plan, link consequences to actions, use smiley face charts, daily versus weekly reward systems, individual versus whole class reward systems
	Positive relationships	Importance of teacher-child relationship, positive agenda, positive link, teacher remaining calm, use low voice, soothing, children take cues from teacher, use discrete messages, subtly engage child, serene personality of adults, TA-child relationship, understand what triggers the child and avoid those in the first instance

Three types of problematic behaviours

Speaking with these SENCOs identified three different types of problematic behaviours among children with SEN: 1) noncompliance, 2) avoidance, and 3) aggression. Individual children might demonstrate one or more types of problematic behaviour. The SENCOs explained that when these behaviours were present among children identified with SEN, the behaviours could be more severe in both frequency and intensity than among other children.

For example, one SENCo describes the case of a child in Year 4 identified with combined ASC and ADHD. This child demonstrates noncompliant behaviour regularly in the classroom:

‘When he makes his mind up that he’s not going to do something, no matter how much you try to coax or cajole or whatever, it’s very, very difficult to win, to turn the situation around...If he perceives that the task that’s being set in the class [is something] that he’s done before, for example in his old school where he came from, then there’s no way you can change his mind.’ (SEnCo 1)

Described by the SENCo as ‘higher functioning’, this child clings quite stubbornly to his decisions, and his aversion to directing attention to familiar tasks is partially explained by the SENCo as being amplified given the child’s diagnoses of both ASC and ADHD.

Another behaviour described by SENCos involves children avoiding unfavourable activities or situations by removing themselves from an area without permission or even against instruction. In this following example, one SENCo describes a child identified with SEN who has displayed such problematic behaviours since he first entered school in Reception:

‘Basically, he’s what we sort of term here a “runner and hider”. So, basically if something doesn’t go his way, he will run out of the room and hide in the cloakroom, or he will go under a table.’ (SEnCo 3)

This type of behaviour may closely resemble noncompliance, as avoidant behaviour includes ignoring or refusing to comply with directions. However, avoidance adds an additional component of physically moving away from the adverse activity or person.

Children disrupt their own learning as well as that of their classmates when engaged in noncompliant or avoidant behaviours in school. However, some children also act in aggressive ways, towards themselves or others. As might be expected, such violence in the classroom raises concerns about the safety of all involved:

'[Most] of the children here, if they have a high level of need, they will kick off in a big way...that will lead the teacher to getting the class out of the classroom...leaving the child in there wrecking the classroom...we wouldn't do anything without a member of the Senior Leadership there, and we work in a team, because these children will escalate and act out in very, very violent manners. So, we have to ensure that they're safe and the class is safe as well.' (SENCo 2)

This point about ensuring the safety of the child acting out, as well as other children and adults in the room, is central when dealing with aggressive behaviours. The examples of children acting aggressively often involved escalation of the violent behaviour when adults intervened, indicating that emotionally charged situations such as these frequently became worse before resolving.

When behaviours escalate, the increased risk threatening anyone near the outburst presents an additional challenge. The following example is of a six-year-old boy identified with ASC, whose aggressive behaviours illustrate such a crescendo:

'[He] would spit and hit and bite and kick. As that happened, the person who was working with him would try and remove him from that situation, and then he would just escalate and become more and more anxious, and he would throw chairs and smash a window, and he would just become more and more troubled. Then, he would have to be restrained, and then he really, really hated that. So, that just made it worse, but we had to do that to keep other people safe.' (SENCo 4)

Thus, we see how initial aggressive behaviours are problematic for teachers, but the increased violence that might occur when trying to intervene with children who are acting aggressively may prove even more concerning. As situations deteriorate, emotions heighten for all involved.

The frustration of not knowing how to de-escalate dangerous behaviours, of dealing with problematic behaviours that endure through academic years and despite significant attempts to mediate the situation, and of trying to engage children in their learning while supporting their various educational needs can be great. These educators supporting children identified with SEN and behavioural needs alluded to the importance of determining what lies at the root of the problematic behaviours to successfully respond to them.

Approach to classroom behaviour management: SENCOs first seek to determine a cause

The SENCOs interviewed in this study highlighted the importance they placed on understanding the reason for children's problematic behaviours; that is, before the implementation of any specific strategy, they first sought to determine an underlying cause. Through observing children and assessing their behaviours, these educators work within a team of teachers, school leaders, parents, and other professionals to target the root of the problem through selecting and implementing appropriate management strategies based on each individual situation.

One SENCO describes adapting a tool from the educational psychologist to create a 'Barrier to Learning' checklist as a way of standardising the observation/assessment process:

'If it's low level disruption I would go into class and observe, and I've sort of developed this... "Barrier to Learning" [checklist] ...I've done it five minute blocks throughout class, so I can actually look at different behaviours they are exhibiting every five minutes. I observe them for about an hour, or part of the lesson, and look at their active engagement in learning and also if they are exhibiting any behaviour like fidgeting, talking to others, looking around, being distracted, and [I] look at that over a period of a whole lesson. It's actually been very interesting in that some children can't manage the input, some children can't manage working on their own. So, you can identify key points of the lesson and put in place some kind of strategy the teacher could use, and then monitor that and see if it works.' (SENCo 2)

Using this fixed-interval sampling strategy is incredibly time-consuming, but it provides the SENCo with a wealth of information about how specific children interact within the class. Such in-depth knowledge provides a clearer understanding of children's special educational and behavioural needs, allowing adults to tailor their strategies.

Given the considerable effort by these educators to determine *why* children were behaving in such ways before then developing strategies to reduce/stop the behaviour, it seems worthwhile to ponder whether such attention is necessary. One SENCo referenced Bloom's taxonomy of thinking (Bloom, Engelhart, Furst, Hill, & Krathwohl, 1956) to explain the importance of understanding how educators differentiate their approaches based on children's various SEN:

'[We need] to see if the child was performing at the lower order thinking, or middle order thinking, or higher order thinking, because that's quite a useful thing to know...[for example] we've got one child who's gifted and talented, but he's working off the scale as far as behaviour goes. So, that makes a big difference as far as the intervention goes.'

Whereas, if you've got lower order thinking skills being exhibited, and [the child is] acting off the scale, then a different type of intervention will be needed. So, it's actually sort of paring away what might be causing the problem, and the approaches we use have got to match what the child needs.' (SENCo 2)

The core issue to address, from this perspective, is not necessarily any of the previously described observed behaviours. Rather, these SENCos suggest that the focus needs to be on the underlying cause of any observable behaviours, which could be related to children's special educational needs as well as to other factors, including home life, as one SENCo sums it all up:

'We've found that the more we look at it, the more you've got to pick apart what's actually at the root of it all.' (SENCo 2)

These SENCos advocate for a classroom behaviour management approach that incorporates first observing problematic behaviours and then assessing the situation to identify a cause, before selecting and implementing any of the following types of behaviour management strategies.

Approach to classroom behaviour management: SENCos' strategies

When discussing strategies for managing children's behaviour in school, the SENCos collectively describe a multi-level structure of support that primarily focuses on using a positive approach. The strategies discussed in the interviews and explained in this section encompass school-wide systems, individual classroom techniques, as well as the use of extra supports for those children whose needs are greater than most (see Figure 7.2).

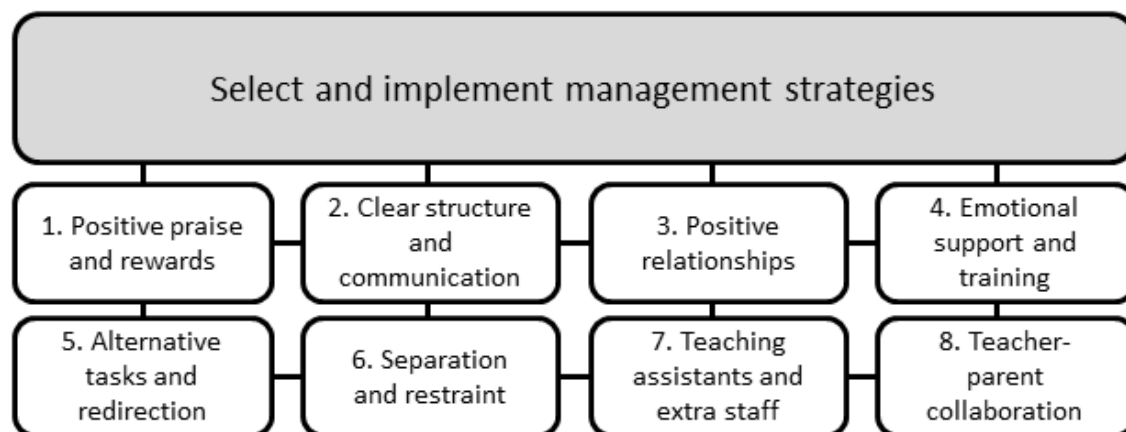


Figure 7.2 Strategies for behaviour management as described by SENCos

Positive praise and rewards. In describing school-wide behaviour systems, the SENCos acknowledge that children do not operate isolated in their classrooms while in school. Instead, they interact with children from other classrooms, older and younger year groups, and they transition to new teachers with each academic year. As such, school-wide consistency is key. Having common strategies in place allows for even very young children as well as those who require extra supports to understand the acceptable boundaries and adopt behavioural norms. SENCos are well placed to advise other teachers and school leaders to implement positive strategies:

‘Well, the way that I view it is that we try to have a positive behaviour approach, both in my classroom and that is what I try to encourage throughout the school when I am advising people. So, I try to put lots of reward systems in place to try and reward good behaviour, and then only sanction poor behaviour when it really is a last resort.’ (SENCo 3)

The SENCos also addressed the use of positive praise as both incentive and reward for children. Praise may be used often in the classroom, but it may also be worked into the whole school environment, as one SENCo describes:

'I do talk a lot about positive praise and finding a way to hook into a child's interest, and also the other point is having a lot of teachers who use key adults in the school to show good work. So, it's actually giving teachers license, if you like, for [acknowledging] "Oh, well, that child's really, really concentrated well on that and...that's been a really good piece of work", so they will send children to me for time out of the classroom—planned movement—and also to get additional praise from a key adult. And likewise, children will go to our head teacher or our deputy head. That's very much in our ethos in school.'
(SENCo 1)

Children are in contact with a variety of adults working in schools, and as such there are reasons beyond establishing a nurturing environment for making sure all staff are well-equipped to similarly respond to behaviours. In situations where children have additional educational and behavioural needs, it is important for any adults who might come into contact with those children to know how to appropriately respond. Take, for instance, the SENCo who is working on communicating with all school adults any additional behavioural needs some children might have:

'We're trying at the moment to create a register of children who need specific strategies, because, of course, children on the playground are different to children who are in the classroom.' (SENCo 2)

Schools may also address supporting the social, emotional, and behavioural needs of children identified with SEN through training staff on inset days:

'Most of the staff are trained in dealing with [autism spectrum]-type social communication issues and identifying how we need to be very literal and visual with children who have these issues. And that is sort of extended to most children, because most of the children in

our school [have a] very high level of need, and there is a very high impact of social deprivation entering school from [their home] environments.’ (SENCo 2)

Here we see how strategies that are meant to support children with very specific social communication difficulties can be uniformly applied for the benefit of all children, regardless of SEN status. Although the strategies highlighted above are all positive, one SENCo also addresses a negative management strategy that she would not allow to occur on a school-wide level:

‘We do have systems in the school for sort of managing behaviour, but what I personally do not like and won’t allow to happen is children with poor behaviour being sent to other people’s classrooms or to younger year groups, because I feel that that is so detrimental to people’s self-esteem that that is more destructive than helpful.’ (SENCo 3)

Thus, there are a number of ways that schools can collectively and positively address behaviour management strategies for all of the children, regardless of any additional educational or behavioural challenges.

Clear structure and communication. The overall message received with regards to classroom management strategies was that it, like the whole school approach, should focus on creating and sustaining ‘a positive, nurturing, caring environment’ (SENCo 6). The SENCos provide a number of ways that teachers may achieve this goal, including communicating clearly, utilizing visual aids and rewards, and developing positive relationships. The following strategies reinforce the previous point that all children can benefit from these positive and proactive strategies, although educators might be more aware of the increased need to implement these techniques when supporting children identified with additional difficulties.

'I always have a very clear structure for [the children with SEN] when they come in, because some of them, you know, will run out and not want to partake. So, I always have a very clear plan, written up, visual, numbered, and the [children]—as we work through it—can see what the beginning is, and I go through very clearly what's expected and what we're learning, and then the children will take it in turns to tick off when we've finished something. So, we can see where we are going, and they know when it's coming to an end.' (SENCo 6)

Here, we see that part of providing clear boundaries for children, particularly those identified with SEN, involves the use of visual props. Children who have difficulties attending to their lessons, need additional reinforcement, or have more severe behavioural concerns all may benefit from the use of visual aids. Sometimes children with SEN who have behavioural needs struggle to understand whole class or long-term rewards, and educators need to accommodate their needs by adapting systems already in place:

'We have a child who is very immature, cognitively, finds making sense of consequences very hard. Hasn't been able to cope with a whole class reward time at the end of the week. So, we've made up a daily system for him. Just through a very, very simple smiley face on a chart or sticker on a chart has helped him – linking it to football, because he's very into football.' (SENCo 7)

Here, we see an example of a whole class reward system that involved a weekly reward, which was insufficient to meet the needs of the child with learning difficulties whose cognitive impairments limited his ability to link rewards (and consequences) to behaviours that did not occur recently. By modifying the existing system to be individual and immediate, the SENCo provides an example of how simple adjustments can make a management strategy more appropriate for children with SEN.

Alternatively, sometimes individual systems put in place for children with SEN who have special behavioural needs can be expanded for use with the whole class:

'[...] he's come to us with ADHD. He's actually being medicated for the ADHD, and obviously our challenge was to maintain his attention, to maintain his focus and concentration so he can join in our mainstream learning. Visual clocks, visual clues so we didn't have to keep telling him things; we could use symbols, or we could point to images to remind him to be sat or to not shout or to be quiet. [...] They are cues specifically designed for this child, but we would use them within the whole class setting. So, it would work for any child in the class as well. But, they are there with this particular child in mind, but we could use them on anybody in the class if we needed to as well, so that child's not feeling singled out.' (SENCo 11)

Therefore, strategies may be modified for more targeted use or more general use as needed in the classroom and to ensure that communication not only fits the children's needs but also to ensure that all children feel comfortable participating and are not unduly singled out. The awareness of how children with SEN and special behavioural needs fit into the classroom illustrates how important teacher-child and child-child relationships are for participation and growth.

Positive relationships. Developing positive relationships and integrating all children, particularly the ones with SEN and behavioural needs, is important for behaviour management. Positive relationships might not come easily to some children identified with SEN who struggle to communicate or otherwise act in a socially acceptable manner. In these situations, teachers need to make the extra effort to establish that positive rapport:

'In my experience, the strength of the relationship that's built up between a class teacher and pupil is paramount... You can see those teachers who actually find a way forward with

a child—they find some positive link working on positive agendas. If they manage to establish that [positive relationship] with a child...that's really important.' (SENCo 1)

Thus, the idea is that a strong relationship based on respect and trust between a class teacher and each child allows for even those vulnerable children to feel safe and experience success in the classroom. Additionally important, particularly when supporting children identified with SEN who struggle behaviourally, is the demeanour of the class teacher in emotionally charged situations:

'Maintaining the voice level at a very, very low level—the calmness of talking even when you are giving directions and giving expectations of how you would like the child to behave—it's all done at a very, very calm, low level. Because, you know, if you get a situation where your voice goes up and you let that child know that...the teacher is getting stressed, then ooh, that child can hook into that and that can actually cause a response that can be quite a manipulative response, and therefore the child might, you know, draw on that or milk that situation. But you know the very successful teachers with regards to behaviour management are those who have a very, very calm delivery with lots of discrete messages, very subtle ways of engaging the child.' (SENCo 1)

However, it is not solely the class teacher who needs to keep a calm disposition. Many children with SEN who have behavioural needs work closely with other professionals during school, and teaching assistants might be assigned to support specific children. One SENCo explains how the calm temperament of a new teaching assistant resulted in immediate change for one young boy diagnosed with ASC:

'His [new] one-to-one, she has a different personality. So, she's very, very calm, and she keeps a distance from this little boy, and she talks very clearly and precisely and concisely. She gives him space, because he's got some sensory needs as well, and she's just very

serene. And he made a comment earlier this week, which was, “What can I do to wind you up?” Obviously, he’d done just that with his other TAs, and he’s got a result from that. Whereas, she is very, very hard [to wind up], she doesn’t show her anxiety level at all. So, yeah, his comment was what can I do to wind you up, and she said, “Nothing. You can’t do anything to wind me up. We’re here together, and we’re going to work a way through this.” And it’s working really well.’ (SENCo 4)

This example demonstrates the crucial role of establishing a connection between adult and child in the context of the classroom and demonstrates how children get their emotional cues from the adults. The staff believed that the boy in the last example had been able to emotionally control others through his outbursts, which increased the anxiety in the class teacher and teaching assistant. However, the new teaching assistant communicates through her actions, speech, and demeanour that she can remain calm even if he cannot. She also acknowledges the sensory needs unique to his ASC by adapting her interactions with him accordingly.

Developing positive relationships involves tuning into what makes each child ‘tick’ and situating their behaviours in the contexts in which they occur:

‘I think the most important thing really is to develop that relationship with children so that we understand what the triggers are, and so that we take steps to try and avoid the triggers in the first place. Also, to understand why the child is behaving in a negative way when they do.’ (SENCo 13)

Positive relationships require strong communication and these factors are particularly necessary when children with SEN requires physical assistance in functioning with daily activities and self-care. As one SENCo describes:

'Recently I've been working with our TAs and the class teacher of a child who has communication and severe learning difficulties. It's a child who also has physical difficulties, and therefore needs some intimate care. Occasionally the child will not be compliant with the changing of a nappy, or when they're soiled. So, helping the staff to make sure that the communication is good on both sides, that the child does understand what is happening – so they are part of the conversation so that we try to address whether it was coming from frustration or why the behaviour was coming across.' (SENCo 7)

Here, the child required adult support for self-care, and frustration on the child's part could result in challenging behaviours. However, through maintaining strong communication with the child to keep positive relationship between all adults (including TAs and specialist occupational therapists) and the child, the SENCo explains how instances of problematic behaviours were either prevented or managed.

Emotional support and training. Indeed, teachers recognise that often children with SEN require even more individualised support for their emotional and behavioural development. Children with such heightened levels of need may require emotional support and training as well as alternative academic tasks. These children might even need to receive this support outside of the classroom or with the help of a teaching assistant working one-on-one with them.

The SENCos often referred to challenges facing children identified with SEN and behavioural needs, particularly with regards to understanding and regulating their feelings. We have already seen how emotionally charged these situations may be for any adults involved. Consequently, it is not surprising that children are similarly impacted. A major difference to note is that children, and particularly those children identified with SEN, might know neither how to comprehend and express their feelings nor how to regain a sense of emotional equilibrium. As one SENCo explains:

'[We use] an approach that looks at the child's emotional well-being through creativity and play. So, children who are really struggling behaviourally, we use a particular member of staff to work with that child [...] to build a relationship with the child and develop their emotional learning through creativity, through play activities and role playing and you explore. So, you develop their emotional understanding of themselves and people around them, and we find that that then positively impacts on their ability to access academic achievement.' (SENCo 10)

The goal of this emotional regulation training is to help children become aware of their emotions so that they can assert more self-control over their actions in fraught situations. One SENCo explains:

'So, for him to be able to understand the feelings that he gets before he becomes violent. He's a very clever little boy, so he understands his feelings, he can put a name to those feelings now, and he can say, "I'm starting to feel really, really anxious. My fists are getting clenched. I can feel myself getting cross. I need to go and—", and then he'll go off and do something different.' (SENCo 4)

SENcos might also refer children with increased levels of emotional and behavioural difficulties to outside professionals who will come into the school to work with the children one-to-one:

'We also have a lot of learning mentors, therapists that work in school...and we can refer for learning mentors to work with children over a period of time in anger management—in controlling reactions in different situations. We also have [the option of referring to an] art therapist, drama therapist, we've got counsellors who work on specific referrals [and] we work very closely with CAMHS [Child and Adolescent Mental Health Services], who do a very similar job if the need is more permanent.' (SENCo 2)

Getting to the point where children can identify and regulate their emotions takes time and practice. Until that time, it is often on the adults in the school to carefully monitor children's emotions and step in to act before outbursts occur.

Alternative tasks and redirection. Another method used to prevent outbursts or redirect children is to assign an alternative task to engage the target child. One SENCo describes working with a class teacher to provide an alternative task to a child diagnosed with ASC, who is extremely noncompliant in specific classroom-based situations:

'What we have done is we've created a folder of activities, which are very carefully chosen, to promote independent learning, and if it is that he's refused to do something, from the classwork, he then has a choice to have a look at the tasks that are set in this other folder, and I would say that since he's been with us, he's actually completed ooh, more than 75% of the tasks set there, as a quick guestimate.' (SEnCo 1)

This SENCo explains the rationale behind offering such an option:

'So, you know, it's actually giving him an alternative task to do when you know that you're not going to win with the other one. It's not a resolution to the fact that he's digging his heels in and he's getting his own way with not doing the task that's set, but he's very rigid in his thinking...So, it's just been about alternative means of actually engaging him in learning so that actually he's doing something rather than going off on his own agenda, and you know, that's been fairly successful, but he's quite an extreme case.' (SEnCo 1)

Another SENCo describes a child with SEN whose academic levels are significantly below her peers and whose avoidant behaviours regularly result in the educators redirecting her to alternative tasks:

'So, if she does join the other children and sits up that's fine, but if she is clearly giving the message that she is not in the right place and she's refusing to come to the carpet and she

200

won't engage, we have a list of planned activities for her to do, which are more in line with a younger, Foundation Stage curriculum. So, she's got some outside play activities, she's got some practical area activities with sort of painting and playdough, those types of things, and then she has more table task activities, and if she is not able to conform with the adult direction and join the rest of the children she is encouraged to go to one of her activities, and we allow her a bit more free flow there. So, she just goes from one thing to another, and we just have to play it a little by ear at the moment.' (SENCo 8)

Thus, there is a certain practicality associated when working with children identified with SEN who have complex needs. SENCos and teachers recognise that they might need to make compromises as they work to support children's emotional and behavioural development. By creating meaningful alternative tasks, they can divert situations that might otherwise escalate into outbursts. Alternative tasks provide children with either added structure or controlled choices, which can be tailored by the SENCos and teachers to align to core lessons that the children need to learn.

Separation and restraint to ensure safety. Sometimes, despite preventative measures, situations intensify to dangerous levels, and management strategies must prioritise the safety and de-escalation of aggressive outbursts. To accomplish both of these goals, school staff may find themselves physically intervening by either restraining children who are acting violently or removing children to another space. Of course, schools only want to use such methods as a last resort and in situations where safety is endangered:

'The whole point is verbal de-escalation as much as possible, and then we taught very simple physical interventions when they are physically acting out.' (SENCo 2)

Certainly, these aggressive scenarios are on the severe end of the spectrum of problematic behaviours in school, but it is also the case that for some children such intense episodes are not infrequent. One SENCo describes the case of a child identified with SEN whose aggressive outbursts present a particular challenge for those working closely with him:

'We've got a particular, quite extreme case of a child that we've initially had to be restraining on a virtually daily basis. We also had to then give him strategies to be able to begin to understand his anger and also strategies for him to be able to regulate that anger, and also we had to put into place literally, I mean physically, a space for him to be able to go to so that all of these things could be managed, and he can be supported.' (SENCo 5)

Responding to such extreme behaviour, which endangers everyone, including the upset child, other children in the class, and the intervening school staff, must be planned to ensure appropriate safety measures are taken. Additionally, schools must continuously evaluate how situations evolve over time and with intervention, modifying their practices accordingly:

'The space has been a moving space, if you like, and you know, initially of course it would be away from the classroom. So, very initially it would be into the head teacher's office or into...a nurture class area that I work in. So, it was into either of those spaces, and then as we were beginning to try and de-escalate things, we were able to create a space just beside his classroom and also, within that, a particular place for him to be able to go to within his own classroom as well.' (SENCo 5)

Thus, we see that the purpose of removing the child acting violently is not to serve as a punishment for the aggressive behaviour. Rather, it is meant as a safety measure for all involved and to provide a change in environment to allow the child to calm down. The

202

intention on the part of the school staff is to try and keep the child as close to his class as possible:

'As he's been able and the staff are all feeling more confident and able and the sort of kicking off, if you like, is not so extreme, and we're not so worried about the other children as well—because that's a big issue with behavioural problems, is you have to be able to consider the other children in the classroom as well—and staff...we've been able to get [the space set aside for this child] closer to the classroom...the idea was that yes, to move it so that it is as close to everything that the rest of his peers are doing.' (SENCo 5)

These situations can be stressful for the adults involved, as well as for the children. As such, some fluidity is required among the school staff when physical separation or restraint is required to keep people safe.

'The only time we actually have to intervene majorly is if she is hurting herself or other children, which probably at the moment is happening maybe three or four times a week, when we are actually having to intervene to prevent injury to others and to keep her safe. [...] It's quite full-on in terms of one adult working with her, so we're trying to set up a flexible arrangement where one adult works with her in the morning and a different adult from the same area is working with her in the afternoon and as or when needed our teacher takes over giving the assistants a break. So, we are just trying to consider the adults' needs as well as the child.' (SENCo 8)

As this last example notes, these particularly extreme cases require multiple adults either at once or in succession to work with the child. The presence of various support staff, including teaching assistants, is in itself a strategy that schools use for behaviour management.

Teaching assistants as targeted support. Among the variety of strategies offered by the SENCOs interviewed here, the SENCOs made particular note to discuss the central role of teaching assistants (TAs) in supporting children identified with SEN who have special behavioural needs. SENCOs assign TAs to work with children identified with SEN, and the TAs may be there to support a child in class, during unstructured playtime, before school, and during transitions:

'So, we have had to employ an extra member of staff to support that child, and that member of staff makes a safe area for him. She guides him in play times and in school in the morning. She is always there for him. We are trying to wean her off of him a little bit, but she always makes sure that she checks in on him at the start of each session. So, even if it's just a look over to him and a thumbs up to say, you know, "I'm here if you need me" and he can look back and respond. But, she's made a safe space for him and again worked hard on trying to talk through his emotions and being able to get him to self-regulate.' (SENCO 4)

TAs are assigned to work with children with a variety of issues and thus must be familiar with the different types of needs children may have (i.e., behavioural, emotional, speech and language), how those needs interact with each other, and which needs have been determined to lie at the root of any observable problems, so that they can appropriately target their support. Such responsibilities speak to requiring considerable training and experience in special needs, and it is clear that the SENCOs, who arrange for some of the TA training, hold TAs in high regard:

'We've got an absolutely brilliant resource, which is our TA who is trained in mental health, and she does lots of, well, we call it SEAL [social and emotional aspects of learning] work, but she'll email teachers and say, 'Right, I'm going to do a few weeks on

anger. Have you got any kids with that? I'm going to do a few weeks on self-esteem or anxiety' whatever. So, she'll do small group work with children who need more of that. She also has about 16 children on her books who have one-to-one sessions with her' (SENCo 9)

When implementing behaviour management strategies with children identified with SEN and special behavioural needs, TAs play a central role at all stages (assessment, intervention, evaluation) and are in the unique position of developing close bonds with the children they are assigned to support.

Teacher-parent collaboration. Finally, classroom behaviour management strategies are not limited solely to the educators within the school. Rather, they extend also to the home and include teacher-parent collaboration and communication. Generally, this was described as a school-wide approach, although the amount or frequency of teacher-parent collaboration would fluctuate depending on need.

'Yes, as a policy at our school, we do try to involve the parents as much as [possible...] it kind of does depend on each particular child. Normally parent involvement is great, because it is good to have their point-of-view, so how they see [their child's] behaviours and their learning from a home point-of-view, because then we can also offer them support for them. So the support is not just in school, but there is something when they go home. It continues through school and at home as well. Lots of times our particular families need support with structure and routines and to manage behaviours as well. We find lots of times once they are in place, the children's learning improves and their needs improves, and they progress a lot quicker. So we really like to have parents on board, but obviously, we can't control how much input the parents put into the child at home.' (SENCo 11)

So, it is not only that the school aims to be connected with parents on a communication level, but also to establish a partnership or collaboration between school and home environments. In this way, the educators and parents can apply similar strategies across the child's environments and send consistent messages to the child.

'We work very closely with families, particularly with children who are struggling behaviourally. When we do our [emotional] work we are very involved with the parents, and we talk a lot with them. When we carry out the [emotional] assessments we work closely with the parents and try and – if they are receptive to it – we will give them activities to do at home, as well. So, there are very close links between home and school. We find that works most effectively. If we can't do that, if the child's parents aren't for a number of reasons as involved with us, then we do what we can in school.' (SENCo 10)

Even though it is not always possible to engage the parents, the SENCos were clear in describing how parent involvement and strong teacher-parent collaboration was the ideal. In part, this was explained in an 'it takes a village' manner. The idea is not only for there to be consistency across school and home environments, but also that the more people brainstorming strategies and involved in the conversation, the better results:

'As a whole we've got much better ideas together than me alone. So, I would be saying, 'Well, let's get the parent in. Let's get the child in. Let's get you [the teacher] in. Is there anyone else that you think has some skills in this – who's got a child with behaviour problems in their class, would you want to go and ask them? [...] It's more of a whole conversation.' (SENCo 15)

Through this approach, the SENCos recognise the key role that parents play in children's development and how the home situation influences the school situation (and vice versa). By collaborating with parents, educators gain a more comprehensive

understanding of the child, have more people generating ideas of how to best support the child, and establish routines and communication techniques that bridge the child's different milieus for anticipated better gains in behaviour and learning outcomes.

SENCo strategies in relation to IY TCM strategies

The strategies listed by the SENCos have been mapped alongside the strategies included on the IY TCM training programme for comparison (see Table 7.2). This method has been used before in determining whether IY TCM content would be acceptable and appropriate when transported to different cultural contexts (Baker-Henningham, 2011). Strategies elicited from the SENCos aligned with all six content areas from the IY TCM Programme, with many concentrating under the IY TCM area of 'building positive relationships with children', which is the foundation of the IY TCM approach. One strategy mentioned by participating SENCos that did not easily map onto the IY TCM strategies involved those extreme situations during which children severely and aggressively acted out, leading educators in the interest of safety to intervene by physically restraining the child. This strategy was implemented not as a consequence but as a measure of security for all given the gravity of the aggressive outburst. Thus, it is not included in those strategies aimed at decreasing inappropriate behaviours. While intentional use of TAs was included under the 'building positive relationships with children' strategy because the SENCos described the positive relationships that the TAs built up with children, it should also be noted that the presence of additional adults in the classroom or school is not a given in all contexts, and reliance on these extra members of staff could be seen as a separate strategy altogether. In these ways, the IY TCM strategies appear to be consistent with the strategies generated from the SENCos. Where the SENCos' strategies did not map on to the IY TCM content (physical intervention, additional adults) these strategies can be seen as an addendum to

ensure everyone's safety and to respond to children whose needs are particularly complex, rather than as being incompatible with the fundamental positive principles guiding IY TCM.

Table 7.2 SENCo-generated strategies mapped alongside IY TCM strategies

Strategies generated by the SENCos	Strategies included in the IY TCM Programme
• Calm demeanour • Collaboration with parents • Developing positive teacher-child relationships • Positive, nurturing, caring environment • Role of support staff (e.g., teaching assistants) • Whole school continuity/consistency	• Building positive relationships with children (e.g., by listening, talking and playing with children, paying attention to and involving children, collaborating with parents)
• Praise from key adults (e.g., Head Teacher, SENCo)	• Teacher attention to positive behaviours, encouragement, and praise
• Rewards for good behaviour or completing lesson objectives (e.g., singing favourite song)	• Incentives as motivation
• Clear lesson structure (i.e., numbered, tick off items as completed) • Individual systems (charts/timetables) • Literal communication (particularly for children identified with ASC) • Visual props	• Proactively preventing problems
• Alternative tasks • Calm down area (inside or outside of classroom) • Verbal de-escalation	• Decreasing inappropriate behaviour through ignoring, redirecting, and consequences
• Emotional support and training for children (e.g., Thrive) • Strategies to understand and regulate feelings (e.g., anger)	• Building children's social skills and problem-solving competences

Discussion

The experiences and views of the participating SENCos in this study indicate that supporting children identified with SEN who also have behavioural needs requires the use of positive, proactive, and clear behaviour management strategies delivered consistently within the context of a nurturing relationship. The SENCos described dealing with essentially the same types of problematic behaviours among children identified with SEN as those without (e.g., noncompliance, avoidance, aggression). However, the complex needs of these children identified with SEN often meant that the behaviours presented in ways that were perceived to be more disruptive or which persisted across time. The SENCos discussed the need for clear communication (e.g., literal, visual) and constant

attention to the teacher-child relationship to support children's self-esteem, which could be particularly vulnerable due to their identified SEN.

Children spend substantial time in school, and comparisons are made between attachment theory's original caregiver-child focus (Bowlby, 1969) and the importance of a positive teacher-child relationship (C. Murray & Greenberg, 2006; Pianta, Hamre, & Stuhlman, 2003). Children with poorer relationships to their teachers are more likely to present with mental health issues (Lang, Marlow, Goodman, Meltzer, & Ford, 2013; C. Murray & Murray, 2004). This is of particular concern for children identified with SEN, because research suggests that the global attachment orientation of children identified with learning disabilities is less likely to be classified as secure than that of their peers without identified learning disabilities (Al-Yagon & Mikulincer, 2004). According to a review of educators' perceptions, teachers serve as a more critical influence for children identified with SEN than for children without identified SEN (Armstrong, 2014). Yet, teachers report feeling unprepared for inclusion settings after completing their initial teacher education, and classroom behaviour management tops their list of areas in which additional support/instruction is required (Winter, 2006). Armstrong (2014) also shows that professional development opportunities can positively influence teachers' perceptions of children identified with SEN. Given the international movement towards inclusion of children with all types of identified SEN in mainstream schools (Ainscow, Dyson, & Weiner, 2014; UNESCO, 1994), school- or teacher-based interventions might support the prevention of problems as children enter adolescence. Providing professional development in classroom behaviour management is thought to transition teachers from a discipline-oriented approach to one that emphasises positive behaviours and proactive strategies, which subsequently might improve children's well-being and academic outcomes. Indeed, influencing teachers' management strategies towards positive techniques results in fewer

school suspensions and higher academic engagement (Burke, Oats, Ringle, Fichtner, & DelGaudio, 2011).

The comparison of the strategies identified by the SENCOs to those taught on the IY TCM training course suggests that the IY TCM strategies would be acceptable and appropriate for working with these multifaceted children. However, the SENCOs explained that with the heightened severity and complexity of needs presented by children identified with both SEN and behavioural difficulties, there are times when safety must take precedence and their strategies have more to do with de-escalating a dangerous situation than with preventing misbehaviour or delivering a consequence. Furthermore, for these children on the more severe end of the spectrum, additional funding is often needed to hire TAs to work as one-to-one aides. While TAs are able to implement any of the behaviour management strategies, the assignment of a TA to a given child could be viewed as a strategy in its own right. Of course, provision for additional key adults is not part of the IY TCM model, which seeks to address the strategies of the class teacher (not reallocate resources that schools may or may not have). In this way, the IY TCM programme appears generally appropriate for educators supporting children identified with SEN who have behavioural needs even though there may yet be situations where individual cases require alternative responses, such as equipping educators to respond to extreme cases of physical aggression or on how to collaborate with a TA. These complex situations might be rarer than incidences of noncompliance or avoidance, but such exceptional outbursts are concerning to SENCOs in making sure their schools remain safe, nurturing environments for all children.

A strength of this study is that it is the first to explicitly examine the appropriateness of the IY TCM programme for children identified with SEN and who have behavioural difficulties. While other studies have focused on the effectiveness of IY TCM

210

among both clinical populations (Webster-Stratton, Reid, & Hammond, 2004) and mainstream classrooms (Baker-Henningham et al., 2012; Hutchings et al., 2013; Reinke, Stormont, Herman, Wang, et al., 2014), no study has yet focused on the appropriateness of the IY TCM programme for children identified with SEN. By using telephone interviews, this study was able to elicit the views of SENCOs across a wide geographical area and during their busy weekly schedules as the school year came to a close. The SENCOs in this study were, on average, a very experienced group of teachers and therefore the responses provided in this study can be seen as stemming from their considerable experience as educators.

The two-staged analysis process of first thematically analysing data from Round One and then translating the data from round two into the analysis is also a strength of this study. Round One only included six SENCOs, which is enough for overarching themes to emerge but is still usually too few participants to achieve saturation (Guest et al., 2006). It takes a sample of 12 participants for analyses aimed to describe common behaviours/beliefs to sufficiently achieve saturation (Guest et al., 2006). Indeed, after Round Two and the analysis of all 15 participating SENCOs, the confidence that the findings had been sufficiently comprehensive was much increased.

Although this study was intended as exploratory to assess whether IY TCM would be an acceptable and appropriate programme for educators teaching children with both SEN and special behavioural needs, one limitation of this study is that its design does not allow for any conclusions to be made about the actual effectiveness of the IY TCM programme for these children. Given that UK educational policy allows for special provisions for children identified with SEN in school, future research is needed to determine exactly what constitutes 'treatment-as-usual' for these children as well as any effects of the IY TCM programme strategies on children identified with SEN who also

have behavioural difficulties. However, it can be concluded from this study that IY TCM strategies would be acceptable and appropriate (if not sufficient) for supporting children with SEN in mainstream school settings.

DISCUSSION

Chapter 8 : Discussion

Introduction

The two studies that make up this thesis are aimed at understanding the effectiveness of the Incredible Years Teacher Classroom Management (IY TCM) programme, how people experience the IY TCM programme, and whether IY TCM strategies are appropriate when supporting children identified with special educational needs (SEN) who also have special behavioural needs. As discussed in Chapter One, this research project is situated within an evidence-based social intervention approach, which applies mixed methods as it values both external evidence of quantitative effectiveness and qualitative professional expertise. Craig et al. (2008) highlight the long process involved in developing, evaluating, and implementing complex social interventions in their Medical Research Council's (MRC) framework. Using this framework, Study One is located within the evaluation stage, where we evaluate effectiveness and understand change mechanisms, and the implementation stage, where we monitor dissemination of the programme (Craig et al., 2008). To do so, Study One aggregated effect sizes for an overall estimate of effectiveness and interpreted qualitative data to produce a mid-range theory about how people experience participating on the course and implementing it within their practice. Study Two developed out of the findings from Study One and is positioned within the development stage of the MRC framework as it explores IY TCM with relation to a new and specific subgroup of the target population. Together, these studies demonstrate the iterative nature of intervention research, such that the systematic review is not seen solely as an end point of 'what works' but also as a springboard for expanding understanding about 'for whom' and 'how'.

This chapter summarises the findings from Study One and Study Two, discusses conflicts of interest and reflexivity, and highlights both strengths and limitations of the overall research project. It then readdresses the novel contributions of these studies to research before delving into the implications for research and practice, outlining next steps for agenda setting in this field.

Summary of main findings

Study One. This mixed methods systematic review of the IY TCM programme contained nine studies that were conducted in England, Ireland, Jamaica, the United States, and Wales. Five studies comprised the quantitative strand, which systematically reviewed effectiveness evidence and applied a multilevel model to meta-analyse intervention effect sizes on teachers' management strategies and children's problematic behaviours and prosocial skills. Five studies also made up the qualitative strand, which systematically reviewed and meta-synthesised core themes and relationships from interview and focus group data, using a grounded theory approach. One mixed methods study was included in both the quantitative and qualitative strands of the review. After results from both strands were reached, findings from each strand were charted into a series of integrative grids, thus mapping the available quantitative and qualitative evidence in relation to one another. This cross-synthesis allowed for discussion not only about congruence and incongruence of results across strands, but also about the direction of findings, strength of results, and gaps in evidence.

Multilevel meta-analysis. Overall, results of the multilevel meta-analysis supported the effectiveness of IY TCM for both teacher and child outcomes. Specifically, teachers' use of negative classroom management strategies decreased, and their use of positive classroom management strategies increased. Teacher-parent collaboration was

not meta-analysed due to lack of coverage across included studies. Only one study (Rhode Island) reported on teacher-parent collaboration and these data were reported in a conference poster with correspondingly limited information/discussion. Children also benefited from IY TCM intervention in the areas of conduct problems and prosocial behaviours.

All included trials conducted baseline assessments of children's behaviour difficulties. One trial (Kingston (2)) only included those children who were classified as at high-risk for behaviour problems. The other included studies selected a cross-section of low-, medium-, and high-risk children in their data collection and analyses, and they reported intervention effect sizes for the high-risk children as subgroup analyses. The multilevel meta-analysis revealed that children who at baseline were assessed to be at high-risk for behavioural problems tended to experience effect sizes three times greater in magnitude than when evaluating a cross-section of children across low-, medium-, and high-risk baseline behaviour scores.

Children's academic readiness was not meta-analysed, again due to lack of coverage across included studies. Only one study (North Carolina) reported on academic outcomes, and these were reported in a conference paper using preliminary trial data. As such, no conclusions could be posited base on these initial findings.

In some cases, the measure of conduct problems included items that reflected features of hyperactivity/inattention and peer problems. As a result, a *post hoc* multilevel meta-analysis evaluating IY TCM on children's broader behavioural difficulties (e.g., conduct problems, hyperactivity, and peer problems) found statistically significant effects in favour of IY TCM over control, with high-risk children again experiencing benefits three times greater than a cross-section of children.

Due to their aggregative nature, meta-analyses can eventually reach a point at which there is ‘informational redundancy’. At this point it is considered that the overall estimate is established such that there is no statistical doubt, meaning that further studies are unlikely to reverse the result (Booth, Sutton, & Papaioannou, 2016). Due to the small number of included studies in this meta-analysis, reaching informational redundancy was very unlikely. Even so, this multilevel meta-analysis established clear trends in the child-level data whereby effect sizes across all outcomes for high-risk children were three times greater in magnitude than effect sizes for a cross-section of classroom children representing low-, medium-, and high-risk children. This trend is important because it indicates that the desired population (high-risk children) is benefiting from this intervention. Furthermore, this finding is noteworthy because as an intervention, IY TCM does not directly work with children (high-risk or otherwise). Rather, facilitators deliver IY TCM to teachers who then must implement the strategies in their classrooms to reach the children. Because of the layers of implementation inherent in IY TCM, it is important to explore how people experience the programme, which was the aim of the qualitative meta-synthesis.

Qualitative meta-synthesis. The qualitative meta-synthesis revealed a cyclical process of primary and secondary learning on the teachers’ parts, which ultimately led to a wide range of benefits to teachers, their classrooms, and the children. In particular, teachers reported perceptions that IY TCM was effective in increasing their knowledge of child development and classroom management, in widening their locus of control, in improving their emotional well-being, and in revitalising their teaching philosophy and practice. Although these qualitatively reported benefits are unlikely to occur in a strictly linear manner, a preliminary process tracing teacher effects through each of these areas is suggested in the findings of the meta-synthesis. This theory is just one of many outputs

from the meta-synthesis that requires testing in future studies to establish the mechanisms by which teacher practice changes during IY TCM training. This could involve further qualitative studies that explore teachers' perceptions of how they benefitted from IY TCM intervention, but it should also involve quantitative studies that measure these outcomes at different time points to establish a temporal ordering of when effects occur and whether certain outcomes (e.g., improved emotional well-being) mediate the effects in observed teacher management strategies.

The perceived child-level outcomes that were qualitatively reported in primary studies included behaviour, social skills, emotional well-being, academics, and generalising to environments outside the classroom. Of these, the first three were most robustly supported across included studies, with the themes of academics and outside the classroom more tentatively included. In contrast to quantitative meta-analysis, which seeks to aggregate findings around an underlying truth, this interpretive meta-synthesis served to broaden the understanding of IY TCM (Booth et al., 2016). As such, while each included study added to the theory resulting from the synthesis, it was not clear beforehand whether the synthesis would converge or diverge. To the extent that each included study intended to add novel findings (often a goal in publications), it was possible that theoretical saturation would not be reached. Certainly, the few included studies meant that this review was unlikely to reach saturation even were convergence a possibility. Interestingly, the findings of the qualitative meta-synthesis suggested that certain themes are clearly present across studies and with additional research might approach saturation, including: all of the teacher outcomes (knowledge, locus of control, emotional well-being, and philosophy/practice) and children's behaviour, social skills, and emotional well-being outcomes. Other themes (notably many of the critiques) possessed considerable within and between study variability. This was particularly evident with the conflicting finding

that some people felt IY TCM should be implemented more broadly across the whole school (rather than in just one classroom or across one year group) for consistency and impact, even though voluntary participation was thought to be preferable to mandated training. One finding that was clear across qualitative studies was that the video vignettes from the American context did not culturally translate to non-American audiences. Even when participants could appreciate the usefulness of the videos in sparking discussion on relevant topics, the surface-level form of these vignettes (i.e., showing scenarios of American teachers with classrooms of American children) presented an added barrier for facilitators implementing IY TCM in other countries.

Cross-synthesis. In mapping the quantitative results from the multilevel meta-analysis alongside the qualitative findings of the meta-synthesis, it was apparent that the desired outcomes of improving teacher positive management strategies, reducing children's problematic behaviours, and increasing children's prosocial skills were central in all studies and led to the same conclusions. However, there were many areas in which the quantitative and qualitative evidence bases did not overlap. Most noticeably, the qualitative findings wove a story of perceived benefit to teacher outcomes that was considerably broader than has been quantitatively measured in RCTs. Whereas trials relied on measuring changes in teachers observable behaviours with regards to classroom management, teachers reported numerous non-visible benefits including their own emotional well-being. It is unclear to what extent these other benefits might mediate the changes in teacher behaviour or even mediate changes in children's outcomes (e.g., does a calmer/less stressed teacher change child behaviour). To determine this, future studies would need to quantitatively measure them as well as establish temporal sequencing.

One of the most robust qualitative findings was the critique of using American video vignettes for training teachers in non-American contexts. While these critiques

centred mainly on questioning the relevance of these vignettes to non-American audiences, this criticism links into the discussion about culturally adapting such components for acceptability and applicability as well as for any subsequent effects of intervention. In only one of the quantitative studies did researchers culturally adapt the video vignettes. It is noteworthy that the quantitative data, which was mostly from non-American contexts, still suggests that the programme is effective, and so it is unlikely that the American video vignettes present a substantial barrier to success when applied in non-American contexts. However, two questions worth asking in future research include: 1) are effect sizes greater when video vignettes are from the same cultural context as the teachers receiving IY TCM? and 2) would culturally adapting the video vignettes improve buy-in among teachers receiving IY TCM in non-American contexts? Although creating new video vignettes in other countries would be a costly initial investment, it has been done in at least one resource-constrained context (Jamaica). Given the positive results of the multilevel meta-analysis combined with the universal critiques of the videos, there is sufficient rationale that such an investment could be worthwhile prior to large-scale take-up of IY TCM in non-American contexts.

Finally, in addition to a broader range of teacher outcomes, the qualitative findings raised questions about possible moderators for child-level outcomes. Specifically, teachers felt that the IY TCM strategies were more appropriate for younger children and were possibly not effective with very complex children, such as those identified with SEN. As none of the quantitative studies has yet explored such subgroups for possible moderator effects, this is another area where the cross-synthesis highlighted gaps in the evidence. It is this particular area that Study Two addressed.

Study Two. Fifteen SENCOs from schools across Devon participated in semi-structured telephone interviews in this study. The purpose in interviewing school SENCOs

was to establish the acceptability and appropriateness of IY TCM for teachers supporting children identified with SEN who also presented with special behavioural needs. The participating SENCos generated a range of management strategies that they used in their own teaching practice or advised other teachers to use when working with children identified with SEN who also had some behavioural problems. When these SENCo-identified strategies were mapped alongside the IY TCM strategies, the results suggested that both sets of strategies overlapped and thus IY TCM would be acceptable and appropriate for this subgroup of children, although due to additional strategies generated by the SENCos, IY TCM might not be sufficient for these children. Notably, concerns over whether IY TCM would be sufficient stemmed from the emphasis placed on using additional adults as part of regular classroom management for children identified with SEN. Sometimes these teaching assistants (TAs) were assigned to support a single child whose needs were so great that they required individualised attention and support (e.g., to assist with transitions). Alternatively, TAs worked with small groups of children who needed to focus on similar targets (e.g., anxiety, anger). Although most of the classroom management strategies were skills or techniques that the class teacher implemented (and indeed which the TAs could implement), the assignment of TAs to children for assisting with social, emotional, and behavioural development in school is in itself a separate strategy. A second SENCo-generated strategy that did not align with IY TCM content was the separation and restraint of children to ensure safety. This strategy was described mainly as a reactive strategy that was not a planned consequence delivered for problematic behaviour but rather a set of actions taken to minimise harm. Although the IY TCM programme aims to minimise the need for reactive strategies (e.g., time out) due to the application of proactive and positive strategies, the SENCos highlighted some ways in which children with SEN might require separation in regular practice, including when

sensory needs were such that the activity in the classroom was too much for the child with SEN and they needed a quiet, secluded area to take a break. In such situations, an IY TCM approach would prefer to see the child's needs closely monitored such that any sensory breaks were more often implemented proactively rather than after an outburst. Either way, there do appear to be reasons for why classroom management for children identified with SEN will at times require different, additional strategies, even though all of the IY TCM strategies would still apply.

Conflicts of interest and reflexivity

Quantitative reviews of effectiveness trials are situated nearer a positivist position than to a constructivist one, and therefore tend to be more concerned about transparently reporting any conflicts of interest than partaking in reflexive practices. This is because they seek a value-free evaluation of whether interventions affect change on desired outcome variables. In contrast, qualitative studies and reviews do not seek to neutralise researchers' positions (Horsburgh, 2003). Rather, researchers in a qualitative tradition are to be reflexive in their research, meaning that they acknowledge the active roles they play in the design, data collection and analysis, and interpretation of the findings (Horsburgh, 2003). The idea is that theories are bound by the context in which they are developed and interpreted, because they are not facts (i.e., underlying truths) but interactions between those involved in the process (Daly, 1997; Guba & Lincoln, 1989; Horsburgh, 2003). Given the mixed methods nature of this dissertation, addressing both conflicts of interest and reflexivity is appropriate.

Conflicts of interest. There are no conflicts of interest to declare with regard to either study in this dissertation. I have no connection (financial or otherwise) to the Incredible Years programme. Furthermore, the results and subsequent interpretation of

effects within the quantitative strand of Study One are based entirely on independent evaluations of IY TCM rather than on developer-led evaluations. Research in psychiatry (Perlis et al., 2005) and bio-medical (Friedman & Richter, 2004) fields has established that studies with conflicts of interest are much more likely to produce results favourable to the intervention. Eisner (2009) explains the two theories for why this might be the case: 1) high fidelity view and 2) cynical view. First, it could be that when the developer is involved in the evaluation the intervention is delivered with higher fidelity, meaning it is implemented according to plan. By creating a scenario where the intervention is as close to ideal as intended, the potential for stronger effects exists. Second, it could be that when the developer is involved in the evaluation decisions are made that systematically favour producing stronger effects (e.g., selection of specific participants, type of statistical analysis). While developer involvement would not be problematic were the intervention intended to always be implemented in the context of having such an association, it becomes problematic when inferences about the external validity of findings are made to contexts where the participation of the developer is not expected. In the case of this systematic review of effectiveness trials, the developer was not involved in any parts of the review's study design, a protocol (Nye et al., 2015) was published with planned statistical analyses listed *a priori*, and all included studies were independent evaluations of the target programme. Therefore, although it is best to be cautious in asserting unbiased results, features of this review strengthen the confidence in results with minimised bias. It is unlikely that the overall summary effect sizes are overestimated due to conflicts of interest in either the primary studies or this review process.

Reflexivity. Recognising that I approach this research from a perspective that is unique to my context, it is important that I address my background as it pertains to the research conducted in both studies of this dissertation. Certainly, my interest in the social,

224

emotional, and behavioural development of children is in part driven by my educational and professional experiences in the fields of psychology and special education. Having worked as a special education teacher for children identified with various disabilities (e.g., emotional-behavioural, mild- to moderate-learning), I have firsthand knowledge of what it is like to be a teacher who feels underprepared to support these aspects of children's development in the school context. Therefore, while I view interventions that effectively address these areas as much needed, I do not feel any particular preference for one specific intervention (e.g., Incredible Years) to be effective because of any allegiance to the programme. If anything, my training in a critical evidence-based approach to social interventions drew my attention more to the areas of weakness and gaps in knowledge, in a desire to better understand how IY TCM might operate. In collecting and analysing the qualitative data for the review in Study One, it is possible that my experience as a teacher who collaborated with parents and specialists as well as who sought to integrate children's voices into decision-making processes (e.g., individualised education plan meetings) made me more aware of the absence of these groups as indirect stakeholders in the primary literature.

With regards to the semi-structured interviews conducted with SENCOs, my previous teaching certainly served as one source of interest in hearing the perspectives of educators who have SEN roles and responsibilities in mainstream schools. Conversely, there existed many structural differences between the context in which I previously taught (i.e., urban, high-poverty, United States) and the context of the SENCOs whom I interviewed (urban and rural, variety of socio-economic situations, Great Britain). These differences were to the extent that I felt could empathise with the day-to-day struggles and triumphs of these educators and yet I still needed the SENCOs to explain many of the services they described to understand what each meant within their context.

In analysing and reporting the qualitative data for both the review and the primary interviews, I sought to ground the findings in the data through using direct quotations and numbering them to anonymously identify the different sources. In this way, readers might disagree with my interpretation but they will be able to follow my audit trail and confirm the transparency of my interpretation as it is supported in the quotations provided to illustrate the narrative (Koch, 2006; Tong et al., 2012).

Strengths

Although studies have previously systematically reviewed RCTs of classroom management programmes – with (Oliver et al., 2011) or without (Whear et al., 2012) meta-analysis – this is the first mixed methods review on a classroom management programme. Pluye and Hong (2014) assert that the mixed methods review ‘combines the power of stories and the power of numbers’ (p. 30), and in doing so, it capitalises on the strengths of these methods while also compensating for the limitations inherent in each approach individually. In bringing together quantitative and qualitative data on a given topic (in this case, the IY TCM programme), the product is meant to be greater than the sum of its parts (Booth et al., 2016).

Regarding its parts, this multilevel meta-analysis serves to more completely aggregate primary study effect sizes than previous meta-analyses by modelling the range of effect sizes nested within these studies. Rather than allowing for researcher judgment to bias the analysis by selecting only one effect size per study or averaging effect sizes within each study without accounting for how those effect sizes are correlated, the results of the multilevel meta-analysis estimate an overall summary effect that takes into consideration within study variance, between study variance, and residual error.

Additionally, as far as I am aware, this is the first qualitative meta-synthesis of experiences of a teacher classroom management programme. In synthesising people's experiences, this review produced mid-range theory that expands the possible range of benefits from IY TCM intervention and suggests possible mediators and moderators. This better understanding of the range of possible benefits is important because although improving teacher practice and children's social and behavioural skills might be key priorities for schools, they are unlikely to be the only needs requiring intervention. If an intervention that is intended to change these core outcomes also affects secondary outcomes (or mediators) that are priority areas for schools (e.g., teacher and child emotional well-being), then this could be a compelling reason for schools to invest in this training. Similarly, if a deeper understanding of 'for whom' the programme works identifies certain subgroups of children who do not respond as well to intervention as the overall average, then schools serving larger proportions of this subgroup may choose to select a different programme.

Indeed the comprehensive picture around both the effectiveness of IY TCM on a range of outcomes as well as the theory of how people experience the programme and the benefits they report serves to highlight both the areas where information has been triangulated between the two strands as well as the areas where one type of research has failed to inform the other.

Following from Study One, Study Two serves as preliminary evidence assessing current school practice in Devon for classroom management with children identified with SEN. In doing so, one of its strengths is including a new voice to the IY TCM body of literature through purposively sampling SENCos. SENCos were selected because they were well-placed to comment on management strategies for a subgroup of children – those identified with SEN – who might experience moderated effects of IY TCM intervention.

Specifically, teachers whose experiences were included in the qualitative strand of the Study One review perceived that children identified with SEN presented more complex needs, and as such they might not respond to IY TCM strategies in the same way that their peers without identified SEN might respond. Indeed, because effectiveness studies produce average group effects for treatment and control conditions, it is possible that hidden within positive overall group effects are subgroups of children who systematically (as opposed to randomly) respond to treatment in a different way. The only subgroup that had been explored in the IY TCM literature to this point was children who at baseline were identified as being at high-risk of conduct problems. Study Two, therefore, lays the preliminary groundwork to expand our understanding of ‘for whom’ IY TCM is effective.

Furthermore, data collection in Study Two was conducted across two cohorts of schools participating in the STARS Study. The process of analysing the first wave of data inductively and creating overarching themes (Nye et al., 2015) before then deductively applying those themes to the second wave of data while still allowing for inductive analysis to produce new themes meant that the findings were expanded and checked for trustworthiness and relevance.

Limitations

Small sample size. The main limitation across all aspects of both studies in this dissertation is small sample size. As discussed in Chapter 2, there is no agreed upon minimum number of studies required for conducting a multilevel meta-analysis. Although theoretically, only two studies are required because of the reasoning underlying the aggregation of data, more data generally means the overall effect size can be estimated with greater precision. With only five studies included in the quantitative strand (and only four contributing to the multilevel meta-analysis) it was not expected that the multilevel

model would be powered sufficiently to establish statistically significant results. Although many of the estimates were statistically significant, it is the clear trend among the child-level data that stands out as being of most use from this analysis with limited data: high-risk children are likely to experience benefits from IY TCM that are three times greater in magnitude than a cross-section of children with varying risk.

The qualitative strand of the review was similarly affected by small sample size. With only five included studies, the results of the meta-synthesis cannot be said to have reached saturation, as is desired in grounded theory approaches (Glaser & Strauss, 1967). Rather, the findings should be considered as overarching themes (Guest et al., 2006), which will need to be further refined as new evidence is produced. However, the qualitative meta-synthesis was limited not only in the few number of studies available but also in the general lack of depth of the included studies. Although each study covered a breadth of topics, the explanations and interpretations often remained at a more descriptive, frequency count level. This lack of analytical abstraction and interpretation meant that the third layer of analysis produced in this meta-synthesis was limited in the complexity of the connections between themes.

Finally, in Study Two's qualitative interviews, the main limitation is that data could only be collected from SENCOs in intervention schools (or control schools that had received their training after coming off of the waitlist). This decision was made with the broader STARS Study research team in response to concerns that asking SENCOs in control schools might change their practice or inadvertently lead the schools to seek out information about IY TCM and contaminate the trial design. It is possible that the participating SENCOs' responses systematically differed from how they might have responded prior to IY TCM training within their school. This limitation, while real, is unlikely to have negatively affected either the data collection or resulting findings because

only one teacher in each school had been trained in IY TCM and in most cases the SENCo was unable to provide any specific information about IY TCM training or strategies. This suggests that the SENCos were largely uninfluenced by IY TCM during the intervention year and, as such, the classroom management strategies they discussed and school services they reported using to support children identified with both SEN and special behavioural needs were based on their years of experience teaching rather than this intervention. Even so, this could be corroborated in future research that interviews SENCos in schools where IY TCM has not been implemented.

What is ‘treatment-as-usual’? Another limitation of Study One is the uncertainty of what ‘treatment-as-usual’ means within and across contexts. Randomly assigning participants to treatment or control in studies that examine whether there is a causal relationship between an intervention (e.g., IY TCM) and an outcome (e.g., children’s behaviour difficulties) is meant to establish equivalent groups on all variables except for the treatment variable. In this way, these effectiveness studies aim to answer the question, ‘What works?’. In reality, this question is incompletely stated, because it leaves out the comparison component. Researchers are not merely asking, ‘What works?’; they are asking, ‘What works compared to our control?’. In evidence-based medicine, it is standard practice to compare a drug (intervention) against a placebo (control). In evidence-based social intervention, it is standard to have a ‘treatment-as-usual’ control, meaning that the comparison group acts as they would normally do were the study not occurring. In educational trials examining teacher training programmes, like IY TCM, this could include control group teachers attending regularly offered professional development training (e.g., inset training) or collaborating with local service providers (e.g., educational psychologists).

In order to understand whether a given intervention ‘worked’, we also must understand the services against which we were comparing it. This is akin to understanding the composition of the placebo pill in a medical trial. Over two decades ago, Golomb (1995) described the placebo paradox, arguing that ‘The foundation of evidence-based medicine is undermined by the absence of evidence that placebos are inert. It is paradoxical that there is no standard of evidence to support the standard of evidence’ (p. 530). In brief summary, Golomb (1995) argued that there were no regulations for placebos, no evidence on the inertness of placebos, and only rare identification of placebo composition in medical trials at the time. As a consequence, the extent to which differences in outcome effects across trials were due to variation in the placebo controls remained unknown. Fifteen years later, peer-reviewed journal articles of randomised placebo-controlled trials of drugs are still regularly published in the top medical journals without regularly reporting placebo compositions (Golomb et al., 2010).

Of course, a placebo in a drug trial is not a direct analogy to treatment-as-usual control groups in social intervention trials. However, the underlying messages from evidence-based medicine do translate to education trials. The CONSORT (Consolidated Standards of Reporting Trials) statement published first in 1996, updated in 2001 and again in 2010, provides direction on how to report RCTs to improve quality and allow research users to accurately appraise the rigour of the methodology and findings of studies (Begg et al., 1996; Moher, Schulz, Altman, & Group, 2001; Schulz, Altman, & Moher, 2010). According to the 2010 CONSORT Checklist, Item 5 directs researchers to report ‘interventions for each group with sufficient details to allow replication, including how and when they were actually administered’ (Schulz et al., 2010, p. 2). Further explanation online clarifies, ‘If the control group is to receive “usual care” it is important to describe thoroughly what that constitutes’ (CONSORT Group). Researchers who overlook

specifying the nature of the treatment-as-usual condition make it more difficult to interpret the results of their studies. Additionally, the comparability of outcomes across studies (as in systematic reviews and meta-analyses) is compromised without sufficient description of the control condition, just as is expected of the intervention condition.

All of this is easier argued for than accomplished, particularly because unlike placebo-drug trials where the placebo is at least the same within a study (even if its composition is unspecified), standard convention for treatment-as-usual within educational studies might not exist. This is especially relevant for educational policies or large-scale interventions, where the programme theory and functions allow for differential execution across contexts (Moore et al., 2015). For example, while all schools might offer training during inset days, the specific topics selected and the depth which they are covered at one school could differ from another school's based on the priorities within their context.

One way to understand and describe treatment-as-usual is to monitor it in a process evaluation. Process evaluations assess the degree to which interventions are implemented as intended (Moore et al., 2015), and in doing so they serve to guard against Type III error, or drawing conclusions about an intervention that has not been adequately implemented (Basch, Sliepcevich, Gold, Duncan, & Kolbe, 1985). Similarly, if researchers were to conduct process evaluations on the control group, they could specify what treatment-as-usual entailed and discuss typical features of implementation fidelity – such as dose and uptake – to further illustrate the comparison condition. Due to the complexity and costs associated with extending process evaluations to the control group, especially when it is not anticipated that the control is a standard set of experiences, further methodological guidance needs to be developed in this area.

The key take away point from this limitation section on treatment-as-usual is that when conducting systematic reviews of effectiveness trials, it is not just the comparability of interventions that should be considered but also the comparability of controls. While treatment-as-usual is unlikely to be the same across studies, until further work is done in understanding and reporting what treatment-as-usual entails, conclusions from reviews will be limited in explaining variation of intervention effects due to differences in controls.

Novel contribution and implications

Study One. The primary innovations of Study One are methodological and theory development. Although systematic reviews of RCTs are well established among researchers as the most robust way of determining effectiveness of interventions, conducting a mixed methods systematic review of RCT and qualitative evidence is a novel approach. Study One serves to illustrate that the three benefits of mixed methods research identified by Rossman and Wilson (1985) – corroboration, elaboration, and initiation – can be realised when applied to research synthesis of an intervention. In approaching the synthesis from both quantitative and qualitative lenses, this review tested theory and discovered new knowledge (Teddle & Sammons, 2010), producing a whole that is potentially of greater use to decision-makers and researchers than either part alone.

Some of the implications from Study One include a theory of processes and outcomes due to IY TCM intervention, that should be tested in future studies. Particularly, the dearth of evidence to support teacher-parent collaboration as an improved outcome should be prioritised given that this is a stated aim of the programme and that teachers are parents' main source for information and assistance when they have concerns about their children (H. Green et al., 2005). In measuring this outcome, researchers should consider

asking both teachers and parents to comment on their collaboration, as a form of triangulation.

Similarly, there is a noticeable absence of evidence on whether IY TCM affects children's academic attainment. Albeit a more distal goal of the programme, improving children's academic attainment is an aim of IY TCM that has not yet been very well explored. Depending on how academics are measured, this could be one outcome that does not impose additional burden on participants or research teams if the regularly collected data on academic performance could be used for such analyses.

Some programme-level implications for future research include testing whether implementation is better at the school-level than at its current teacher-level, i.e. is it more effective to train all staff in IY TCM? Or, is it the newly qualified teachers who benefit most from IY TCM compared to more experienced teachers (and if so would there be any rationale for initial teacher education programmes to include this training)?

Finally, given the overwhelmingly consistent criticism of the video vignettes in non-American contexts, decision-makers at the practice or policy levels might consider the merits of investing in the initial costs to culturally adapt the form of the videos to their context. In resource-rich contexts that anticipate rolling out a programme broadly, this might be a decision that aims to make IY TCM more acceptable (i.e., facilitate buy-in) and appropriate for teachers. However, research on transporting parenting interventions to new contexts suggests that these interventions are just as effective in the new country as they were in the country of origin (Leijten, Melendez-Torres, Knerr, & Gardner, 2016). These findings along with the overall positive trends from Study One's multilevel meta-analysis, call into question whether such an adaptation would lead to increased effectiveness of the programme. In resource-poor contexts, it might not be worth the investment to re-film the

videos so long as the group facilitators are able to expertly lead discussions that pull out the key principles and messages.

Conclusion

This dissertation aimed to examine the role of teacher training in classroom management strategies to support teacher's professional development and children's social, emotional, and behavioural health. Teachers' concerns over problematic behaviours in the classroom have been long documented (Meister & Melnick, 2003; Veenman, 1984; Wickman, 1928), suggesting that this is a chronic problem that has not yet been properly addressed.

Similarly, concerns over children's mental health are relatively stable across contexts and indicate that around 10 per cent of school-age children reach clinical cut-offs for diagnoses (H. Green et al., 2005; Merikangas et al., 2010; J. Murray et al., 2013; Wittchen et al., 2011) and another 20 per cent demonstrate problematic behaviours (J. Murray et al., 2013).

By intervening at the teacher-level to change teachers' classroom management strategies, the aim is to address both the teacher's needs for increased skills in their profession and the children's social, emotional, and behavioural needs. The IY TCM programme is just one example of classroom management interventions designed to do just that. Results from the studies in this dissertation suggest that IY TCM effectively improves teachers' classroom management strategies as well as children's behavioural and prosocial skills.

Furthermore, there is a range of additional perceived outcomes which could also be improved through IY TCM, but which have not yet been quantitatively studied. Similarly, although children at high-risk for problematic behaviours benefit more from IY TCM than those without problematic behaviours, there may be other subgroups of children who differentially respond to intervention. The exploratory work from Study Two suggests that it would be worthwhile evaluating IY TCM effects on children identified with SEN,

although these children might still need some services-as-usual (such as TAs) to be successful in the mainstream school context. Overall, future research on IY TCM should seek to tease out the mechanisms of intervention and illuminate any moderators (e.g., SEN, age, class size) so that practitioners and policy-makers are best informed as to the effects they can expect to see if they were to select IY TCM for use in their area. Future researchers should keep in mind mixed methods systematic reviews when proposing new syntheses for the knowledge and insights they offer above and beyond effectiveness alone.

References

- Ainscow, M., Dyson, A., & Weiner, S. (2014). From exclusion to inclusion: A review of international literature on ways of *responding* to students with special educational needs in schools. *En-clave Pedagógica, 13*, 13-30. doi: 10.1007/s11125-008-9055-0
- Ainsworth, M. D. (1985). Attachments across the life span. *Bulletin of the New York Academy of medicine, 61*, 792-812.
- Ainsworth, M. D., Blehar, M. C., Waters, E., & Wall, S. (1978). *Patterns of attachment: A psychological study of the strange situation*. Hillsdale, NJ: Lawrence Erlbaum.
- Al-Yagon, M., & Mikulincer, M. (2004). Patterns of close relationships and socioemotional and academic adjustment among school-age children with learning disabilities. *Learning Disabilities, 19*, 12-19. doi:10.1111/j.1540-5826.2004.00085.x
- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: American Psychiatric Association.
- Anderson, R. (2008). New MRC guidance on evaluating complex interventions. *BMJ, 337*, a1937. doi:10.1136/bmj.a1655
- Armstrong, D. (2014). Educator perceptions of children who present with social, emotional and behavioural difficulties: A literature review with implications for recent educational policy in England and internationally. *International Journal of Inclusive Education, 18*, 731-745. doi:10.1080/13603116.2013.823245
- Assembly, U. N. G. (2015). Resolution adopted by the General Assembly on 25 September 2015. *Washington: United Nations*.
- Baker-Henningham, H. (2011). Transporting evidence-based interventions across cultures: using focus groups with teachers and parents of pre-school children to inform the

- implementation of the Incredible Years Teacher Training Programme in Jamaica. *Child: Care, Health and Development*, 37, 649-661. doi:10.1111/j.1365-2214.2011.01208.x
- Baker-Henningham, H., Scott, S., Jones, K., & Walker, S. (2012). Reducing child conduct problems and promoting social skills in a middle-income country: cluster randomised controlled trial. *The British Journal of Psychiatry*, 201, 101-108. doi:10.1192/bjp.bp.111.096834
- Baker-Henningham, H., & Walker, S. (2009). A qualitative study of teacher's perceptions of an intervention to prevent conduct problems in Jamaican pre-schools. *Child: Care, Health and Development*, 35, 632-642. doi:10.1111/j.1365-2214.2009.00996.x
- Baker-Henningham, H., Walker, S., Powell, C., & Gardner, J. M. (2009a). A pilot study of the Incredible Years Teacher Training programme and a curriculum unit on social and emotional skills in community pre-schools in Jamaica. *Child: Care, Health and Development*, 35, 624-631. doi:10.1111/j.1365-2214.2009.00964.x
- Baker-Henningham, H., Walker, S., Powell, C., & Gardner, J. M. (2009b). Preventing behaviour problems through a universal intervention in Jamaican basic schools: a pilot study. *The West Indian medical journal*, 58, 460-464.
- Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice-Hall, Inc.
- Bandura, A., & McClelland, D. C. (1977). *Social learning theory*. Englewood Cliffs, NJ: Prentice-Hall, Inc.
- Barbour, R. (2014). *Introducing Qualitative Research: A Student's Guide* (2nd ed.). London: SAGE.

- Basch, C. E., Sliepcevich, E. M., Gold, R. S., Duncan, D. F., & Kolbe, L. J. (1985).
Avoiding type III errors in health education program evaluations: a case study.
Health Education & Behavior, 12(3), 315-331.
- Bateman, I. J., & Jones, A. P. (2003). Contrasting conventional with multi-level modeling
approaches to meta-analysis: expectation consistency in UK woodland recreation
values. *Land Economics*, 79, 235-258. doi: 10.2307/3146869
- Begg, C., Cho, M., Eastwood, S., Horton, R., Moher, D., Olkin, I., . . . Simel, D. (1996).
Improving the quality of reporting of randomized controlled trials: the CONSORT
statement. *Jama*, 276(8), 637-639.
- Bernal, G., Jiménez-Chafey, M. I., & Domenech Rodríguez, M. M. (2009). Cultural
adaptation of treatments: A resource for considering culture in evidence-based
practice. *Professional Psychology: Research and Practice*, 40, 361.
doi:10.1037/a0016401
- Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956).
Taxonomy of educational objectives: The classification of educational goals.
Handbook 1: Cognitive domain. New York: David McKay.
- Bonell, C., Jamal, F., Harden, A., Wells, H., Parry, W., Fletcher, A., . . . Whitehead, M.
(2013). Systematic review of the effects of schools and school environment
interventions on health: evidence mapping and synthesis. *Public Health Research*,
1. doi:10.3310/phr01010
- Booth, A., Sutton, A., & Papaioannou, D. (2016). *Systematic approaches to a successful
literature review*. London, UK: Sage.
- Borenstein, M., Hedges, L. V., Higgins, J., & Rothstein, H. R. (2009). *Introduction to
meta-analysis*. Chichester, UK: John Wiley & Sons Ltd.

- Bowlby, J. (1958). The nature of the child's tie to his mother. *The International journal of psycho-analysis*, 39, 350-373.
- Bowlby, J. (1960). Separation anxiety. *The International journal of psycho-analysis*, 41, 89.
- Bowlby, J. (1969). *Attachment and loss* (Vol. 1). London, UK: Hogarth Press and the Institute of Psycho-Analysis.
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative research in psychology*, 3, 77-101. doi:10.1191/1478088706qp063oa
- Britten, N., Campbell, R., Pope, C., Donovan, J., Morgan, M., & Pill, R. (2002). Using meta ethnography to synthesise qualitative research: A worked example. *Journal of Health Services Research and Policy*, 7(4), 209-215.
doi:10.1258/135581902320432732
- Brophy, J. (2013). History of research on classroom management. In C. M. Evertson & C. S. Weinstein (Eds.), *Handbook of Classroom Management*. London, UK: Routledge.
- Burke, R. V., Oats, R. G., Ringle, J. L., Fichtner, L. O. N., & DelGaudio, M. B. (2011). Implementation of a Classroom Management Program with Urban Elementary Schools in Low-Income Neighborhoods: Does Program Fidelity Affect Student Behavior and Academic Outcomes? *Journal of Education for Students Placed at Risk (JESPAR)*, 16, 201-218. doi:10.1080/10824669.2011.585944
- Chapple, A., & Rogers, A. (1998). Explicit guidelines for qualitative research: a step in the right direction, a defence of the 'soft' option, or a form of sociological imperialism? *Family practice*, 15, 556-561. doi:10.1093/fampra/15.6.556
- Cohn, L. D., & Becker, B. J. (2003). How meta-analysis increases statistical power. *Psychological methods*, 8, 243. doi:10.1037/1082-989X.8.3.243

- Collishaw, S., Maughan, B., Goodman, R., & Pickles, A. (2004). Time trends in adolescent mental health. *Journal of child psychology and psychiatry*, 45, 1350-1362. doi:10.1111/j.1469-7610.2004.00335.x
- Conduct Problems Prevention Research Group (CPPRG). (1995). *Teacher - Social Competence Scale*. Retrieved from <http://www.fasttrackproject.org>
- Conners, C. K. (2001). *Conners' rating scales revised*: Multi-Health Systems, Incorporated.
- CONSORT Group. CONSORT Checklist. Retrieved from <http://www.consort-statement.org/checklists/view/32-consort/78-interventions#tab-explanation>
- Corbin, J., & Strauss, A. (2008). *Basics of qualitative research: Techniques and procedures for developing grounded theory* (3rd ed.). Thousand Oaks, CA: SAGE.
- Côté, L., & Turgeon, J. (2005). Appraising qualitative research articles in medicine and medical education. *Medical teacher*, 27, 71-75. doi:10.1080/01421590400016308
- Craig, P., Dieppe, P., Macintyre, S., Michie, S., Nazareth, I., & Petticrew, M. (2008). Developing and evaluating complex interventions: the new Medical Research Council guidance. *BMJ*, 337. doi:10.1136/bmj.a1655
- Creswell, J. W., & Plano Clark, V. L. (2011). *Designing and Conducting Mixed Methods Research* (2nd ed.). Thousand Oaks, CA: SAGE.
- Daly, K. (1997). Re-placing theory in ethnography: A postmodern view. *Qualitative Inquiry*, 3(3), 343-365.
- Darney, D., Reinke, W. M., Herman, K. C., Stormont, M., & Ialongo, N. S. (2013). Children with co-occurring academic and behavior problems in first grade: Distal outcomes in twelfth grade. *Journal of School Psychology*, 51, 117-128. doi:10.1016/j.jsp.2012.09.005

- Davenport, J., & Tansey, A. (2009). *Outcomes of an Incredible Years Classroom Management Programme with teachers from multiple schools*. Retrieved from Trinity College Dublin, National Educational Psychological Service:
- Department for Education, & Department of Health. (2015). *Special educational needs and disability code of practice: 0 to 25 years: Statutory guidance for organisations which work with and support children and young people who have special educational needs or disabilities*.
- DiLalla, L. F., Marcus, J. L., & Wright-Phillips, M. V. (2004). Longitudinal effects of preschool behavioral styles on early adolescent school performance. *Journal of School Psychology, 42*, 385-401. doi:10.1016/j.jsp.2004.05.002
- Dixon-Woods, M., Shaw, R. L., Agarwal, S., & Smith, J. A. (2004). The problem of appraising qualitative research. *Quality and Safety in Health Care, 13*, 223-225. doi:10.1136/qshc.2003.008714
- Eisner, M. (2009). No effects in independent prevention trials: can we reject the cynical view? *Journal of experimental Criminology, 5*(2), 163-183.
- Elliott, D. S., & Mihalic, S. (2004). Issues in disseminating and replicating effective prevention programs. *Prevention Science, 5*, 47-53. doi:10.1023/B:PREV.0000013981.28071.52
- Evertson, C. M., & Weinstein, C. S. (2013). *Handbook of classroom management: Research, practice, and contemporary issues*. London: Routledge.
- Eyberg, S. M., & Pincus, D. (1999). *Eyberg Child Behavior Inventory & Sutter-Eyberg Student Behavior Inventory-Revised: Professional Manual*. Odessa, FL: Psychological Assessment Resources.
- Eyberg, S. M., & Robinson, E. A. (1981). *Dyadic parent-child interaction coding system*. Retrieved from <http://www4.ujaen.es/~mramos/AsMPIA/dpicsmanual.pdf>

- Fauth, R., Parsons, S., & Platt, L. (2014). *Convergence or divergence?: a longitudinal analysis of behaviour problems among disabled and non-disabled children aged 3 to 7 in England. CLS Working Paper 2014/13*. Retrieved from London: Centre for Longitudinal Studies
- Flemming, K. (2009). Synthesis of quantitative and qualitative research: an example using Critical Interpretive Synthesis. *Journal of Advanced Nursing*, 66, 201-217.
doi:10.1111/j.1365-2648.2009.05173.x
- Ford, T., Edwards, V., Sharkey, S., Ukoumunne, O. C., Byford, S., Norwich, B., & Logan, S. (2012). Supporting teachers and children in schools: the effectiveness and cost-effectiveness of the Incredible Years teacher classroom management programme in primary school children: a cluster randomised controlled trial, with parallel economic and process evaluations. *BMC public health*, 12, 719. doi:10.1186/1471-2458-12-719
- Fraser, M. W., Richman, J. M., Galinsky, M. J., & Day, S. H. (2009). *Intervention research: Developing social programs*. Oxford, UK: Oxford University Press.
- Freeman, R. (2009). Health promotion and the randomised controlled trial: a square peg in a round hole? *BMC oral health*, 9, 1. doi:10.1186/1472-6831-9-1
- Friedman, L. S., & Richter, E. D. (2004). Relationship between conflicts of interest and research results. *Journal of general internal medicine*, 19(1), 51-56.
- Glaser, B., & Strauss, A. (1967). *The discovery grounded theory: strategies for qualitative inquiry*. Hawthorne, NY: Aldine de Gruyter.
- Goldstein, H. (1987). *Multilevel models in educational and social research*. Oxford, UK: Oxford University Press.

- Goldstein, H., Yang, M., Omar, R., Turner, R., & Thompson, S. (2000). Meta - analysis using multilevel models with an application to the study of class size effects. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 49, 399-412. doi:10.1111/1467-9876.00200
- Golomb, B. A. (1995). Paradox of placebo effect. *Nature*, 375(6532), 530.
- Golomb, B. A., Erickson, L. C., Koperski, S., Sack, D., Enkin, M., & Howick, J. (2010). What's in placebos: who knows? Analysis of randomized, controlled trials. *Annals of internal medicine*, 153(8), 532-535.
- Goodman, R. (1997). The Strengths and Difficulties Questionnaire: A research note. *Journal of child psychology and psychiatry*, 38, 581-586. doi:10.1111/j.1469-7610.1997.tb01545.x
- Graue, E., Hatch, K., Rao, K., & Oen, D. (2007). The wisdom of class-size reduction. *American Educational Research Journal*, 44, 670-700. doi:10.3102/0002831207306755
- Green, H., McGinnity, Á., Meltzer, H., Ford, T., & Goodman, R. (2005). *Mental health of children and young people in Great Britain, 2004*: Basingstoke: Palgrave Macmillan.
- Green, J., & Thorogood, N. (2014). *Qualitative methods for health research* (3rd ed.). London, UK: SAGE.
- Greene, J. C., & McClintock, C. (1985). Triangulation in evaluation: Design and analysis issues. *Evaluation Review*, 9, 523-545. doi:10.1177/0193841X8500900501
- Greenhalgh, T., Humphrey, C., Hughes, J., Macfarlane, F., Butler, C., & Pawson, R. (2009). How do you modernize a health service? A realist evaluation of whole -

scale transformation in London. *Milbank Quarterly*, 87, 391-416.

doi:10.1111/j.1468-0009.2009.00562.x

Griner, D., & Smith, T. B. (2006). Culturally adapted mental health intervention: A meta-analytic review. *Psychotherapy: Theory, research, practice, training*, 43, 531-548.
doi:10.1037/0033-3204.43.4.531

Guba, E. G., & Lincoln, Y. S. (1989). *Fourth generation evaluation*: Sage.

Guest, G., Bunce, A., & Johnson, L. (2006). How many interviews are enough? An experiment with data saturation and variability. *Field methods*, 18, 59-82.
doi:10.1177/1525822X05279903

Hansford, L., Sharkey, S., Edwards, V., Ukoumunne, O., Byford, S., Norwich, B., . . . Ford, T. (2015). Understanding influences on teachers' uptake and use of behaviour management strategies within the STARS trial: Process evaluation protocol for a randomised controlled trial. *BMC public health*, 15, 119. doi:10.1186/s12889-015-1486-y

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. London, UK: Academic Press Inc.

Hickey, G., McGilloway, S., Hyland, L., Leckey, Y., Kelly, P., Bywater, T., . . . O'Neill, D. (2015). Exploring the effects of a universal classroom management training programme on teacher and child behaviour: A group randomised controlled trial and cost analysis. *Journal of Early Childhood Research*, 15, 1-21.
doi:10.1177/1476718X15579747

Higgins, J. P., & Green, S. (2008). *Cochrane handbook for systematic reviews of interventions* (Vol. 5): Wiley Online Library.

Higgins, J. P., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *BMJ*, 327, 557-560. doi:10.1136/bmj.327.7414.557

- Horsburgh, D. (2003). Evaluation of qualitative research. *Journal of clinical nursing*, 12(2), 307-312.
- Hutchings, J., Daley, D., Jones, K., Martin, P., Bywater, T., & Gwyn, R. (2007). Early results from developing and researching the Webster-Stratton Incredible Years Teacher Classroom Management Training Programme in North West Wales. *Journal of Children's Services*, 2, 15-26. doi:10.1108/17466660200700023
- Hutchings, J., Martin-Forbes, P., Daley, D., & Williams, M. E. (2013). A randomized controlled trial of the impact of a teacher classroom management program on the classroom behavior of children with and without behavior problems. *Journal of School Psychology*, 51, 571-585. doi:10.1016/j.jsp.2013.08.001
- Hyland, L. (2014). *The Incredible Years Teacher Classroom Management Programme in Ireland: A process evaluation and observational assessment of teacher-pupil outcomes*. (PhD), National University of Ireland Maynooth.
- Kagan, D. M. (1992). Professional growth among preservice and beginning teachers. *Review of educational research*, 62, 129-169. doi:10.3102/00346543062002129
- Kennedy, M. D. (2004). *Implementation of the Incredible Years Model Program in Three Colorado Communities: A Case Study*. Retrieved from <http://search.proquest.com/docview/61561177?accountid=13042>
http://oxfordsfx.hosted.exlibrisgroup.com/oxford?url_ver=Z39.88-2004&rft_val_fmt=info:ofi/fmt:kev:mtx:dissertation&genre=dissertations+%26+theses&sid=ProQ:Social+Services+Abstracts&atitle=&title=Implementation+of+the+Incredible+Years+Model+Program+in+Three+Colorado+Communities%3A+A+Case+Study&issn=&date=2005-02-01&volume=&issue=&spage=&au=Kennedy%2C+Mitzi+D&isbn=&jtitle=&bttitle=&rft_id=info:eric/&rft_id=info:doi/ Social Services Abstracts database.

- Koch, T. (2006). Establishing rigour in qualitative research: the decision trail. *Journal of Advanced Nursing*, 53(1), 91-100.
- Kollins, S. H., Epstein, J. N., & Conners, K. C. (2004). Conners' Rating Scales - Revised. In M. E. Maruish (Ed.), *The use of psychological testing for treatment planning and outcomes assessment* (3rd ed., pp. 215-232): Lawrence Erlbaum Associates.
- Korpershoek, H., Harms, T., de Boer, H., van Kuijk, M., & Doolaard, S. (2014). Effective classroom management strategies and classroom management programs for educational practice. *Groningen: GION onderwijs/onderzoek*.
- Lang, I., Marlow, R., Goodman, R., Meltzer, H., & Ford, T. (2013). Influence of problematic child-teacher relationships on future psychiatric disorder: Population survey with 3-year follow-up. *The British Journal of Psychiatry*, 202, 336-341. doi:10.1192/bjp.bp.112.120741
- Lau, A. S. (2006). Making the case for selective and directed cultural adaptations of evidence - based treatments: Examples from parent training. *Clinical psychology: Science and practice*, 13, 295-310. doi:10.1111/j.1468-2850.2006.00042.x
- Legislation.gov.uk. (2010). *Equality Act 2010*. London, UK: HMSO Retrieved from <http://www.legislation.gov.uk/ukpga/2010/15/contents>.
- Leijten, P., Melendez-Torres, G., Knerr, W., & Gardner, F. (2016). Transported versus homegrown parenting interventions for reducing disruptive child behavior: a multilevel meta-regression study. *Journal of the American Academy of Child & Adolescent Psychiatry*, 55(7), 610-617.
- Maas, C. J., & Hox, J. J. (2005). Sufficient sample sizes for multilevel modeling. *Methodology*, 1, 86-92. doi:10.1027/1614-2241.1.3.86

Marlow, R., Hansford, L., Edwards, V., Ukoumunne, O. C., Norman, S., Ingarfield, S., . . .

Ford, T. (2015). Teaching classroom management-A potential public health intervention? *Health Education, 115*, 230-248. doi:10.1108/HE-03-2014-0030

Martin, P. A. (2009). *From small acorns: The positive impact of adopting simple teacher classroom management strategies on global classroom behaviour and teacher-pupil relationships*. (PhD), University of Wales, Bangor.

Mason, M. (2010). *Sample size and saturation in PhD studies using qualitative interviews*. Paper presented at the Forum qualitative Sozialforschung/Forum: qualitative social research.

McGilloway, S., Hyland, L., Mháille, G., Lodge, A., O'Neill, D., Kelly, P., . . . Donnelly, M. (2010). *Positive classrooms, positive children: A randomised controlled trial to investigate the effectiveness of the incredible years teacher classroom management programme in an Irish context (short-term outcomes)*. Retrieved from Archways Ireland: www.iyirelandstudy.ie

McGilloway, S., Mháille, G., Furlong, M., Hyland, L., Leckey, Y., Kelly, P., . . . Donnelly, M. (2012). *Parents, teachers, and early childhood intervention: Long-term outcomes of the Incredible Years Parent and Teacher Classroom Management training programmes (combined 12-month report)*. Retrieved from Archways Ireland: www.iyirelandstudy.ie

McMahon, R. J., Wells, K. C., & Kotler, J. S. (2006). Conduct problems. In E. J. Mash & R. A. Barkley (Eds.), *Treatment of childhood disorders* (Vol. 3, pp. 137-268). New York, NY: Guilford Press.

Meister, D. G., & Melnick, S. A. (2003). National new teacher study: Beginning teachers' concerns. *Action in Teacher Education, 24*, 87-94.

doi:10.1080/01626620.2003.10463283

- Merikangas, K. R., He, J.-P., Brody, D., Fisher, P. W., Bourdon, K., & Koretz, D. S. (2010). Prevalence and treatment of mental disorders among US children in the 2001–2004 NHANES. *Pediatrics*, *125*, 75-81. doi:10.1542/peds.2008-2598
- Merrell, K. W. (1996). Assessment of social skills and behavior problems in early childhood: the Preschool and Kindergarten Behavior Scales. *Journal of Early Intervention*, *20*, 132-145.
- Mihalic, S. (2002). *The importance of implementation fidelity*. Retrieved from Boulder, CO: <http://incredibleyears.com/wp-content/uploads/fidelity-importance.pdf>
- Moher, D., Schulz, K. F., Altman, D. G., & Group, C. (2001). The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials: Elsevier.
- Moore, G. F., Audrey, S., Barker, M., Bond, L., Bonell, C., Hardeman, W., . . . Wight, D. (2015). Process evaluation of complex interventions: Medical Research Council guidance. *BMJ*, *350*, h1258.
- Murray, C., & Greenberg, M. T. (2006). Examining the importance of social relationships and social contexts in the lives of children with high-incidence disabilities. *The Journal of Special Education*, *39*, 220-233. doi:10.1177/00224669060390040301
- Murray, C., & Murray, K. M. (2004). Child level correlates of teacher-student relationships: An examination of demographic characteristics, academic orientations and behavioral orienations. *Psychology in the Schools*, *41*, 751-762. doi:10.1002/pits.20015
- Murray, D. W. (2009). Effects of classroom management training on early learning skills. North Carolina, USA: Institute of Education Sciences R305A090361.
- Murray, D. W., Murr, N., & Rabiner, D. L. (2012). *Preliminary effects of the Incredible Years Teacher Training Program on classroom management skills*. Paper

presented at the Society for Research on Educational Effectiveness.

<http://search.proquest.com/docview/1011396534?accountid=13042>

Murray, D. W., Rabiner, D. L., & Carrig, M. M. (2014). *Grade level effects of the Incredible Years Teacher Training Program*. Paper presented at the Society for Research on Educational Effectiveness.

Murray, J., Anselmi, L., Gallo, E. A. G., Fleitlich-Bilyk, B., & Bordin, I. A. (2013). Epidemiology of childhood conduct problems in Brazil: systematic review and meta-analysis. *Social psychiatry and psychiatric epidemiology*, 48, 1527-1538. doi:10.1007/s00127-013-0695-x

Nye, E. (2013). *Effectiveness of the Incredible Years Teacher Classroom Management Program on teacher and student outcomes: Systematic review and meta-analysis*. (MSc in Evidence-Based Social Intervention), University of Oxford.

Nye, E., Melendez-Torres, G. J., & Bonell, C. (2016a). Context and implications document for: Origins, methods and advances in qualitative meta-synthesis. *Review of Education*, 4, 80-83. doi:10.1002/rev3.3072

Nye, E., Melendez-Torres, G. J., & Bonell, C. (2016b). Origins, methods and advances in qualitative meta-synthesis. *Review of Education*, 4, 57-79. doi:10.1002/rev3.3065

Nye, E., Melendez-Torres, G. J., & Gardner, F. (2015). *Mixed methods systematic review of a teacher classroom management programme: Effectiveness and stakeholders' experiences*. Retrieved from PROSPERO CRD42015020359: http://www.crd.york.ac.uk/PROSPERO/display_record.asp?ID=CRD42015020359

Oldfield, J., Humphrey, N., & Hebron, J. (2015). Cumulative risk effects for the development of behaviour difficulties in children and adolescents with special educational needs and disabilities. *Research in developmental disabilities*, 41, 66-75. doi:10.1016/j.ridd.2015.05.010

- Oliver, R., Reschly, D., & Wehby, J. (2011). The effects of teachers' classroom management practices on disruptive, or aggressive student behavior: A systematic review. *Campbell Systematic Reviews*, 7. doi:10.4073/csr.2011.4
- Padgett, D. K. (2008). *Qualitative methods in social work research* (Vol. 36): SAGE.
- Parsons, S., & Platt, L. (2013). *Disability among young children: Prevalence, heterogeneity and socio-economic disadvantage. CLS Working Paper 2013/11*. Retrieved from London Centre for Longitudinal Studies
- Patterson, G. R., DeBaryshe, B. D., & Ramsey, E. (1989). A developmental perspective on antisocial behavior. *American Psychologist*, 44, 329-335. doi:10.1037/0003-066X.44.2.329
- Patterson, G. R., Reid, J. B., & Dishion, T. J. (1992). *Antisocial boys* (Vol. 4). Eugene, OR: Castalia Pub Co.
- Pawson, R., & Tilley, N. (1997). *Realistic evaluation*. London, UK: SAGE.
- Perlis, R. H., Perlis, C. S., Wu, Y., Hwang, C., Joseph, M., & Nierenberg, A. A. (2005). Industry sponsorship and financial conflict of interest in the reporting of clinical trials in psychiatry. *American Journal of Psychiatry*, 162(10), 1957-1960.
- Pianta, R. C., Hamre, B. K., & Stuhlman, M. (2003). Relationships between teachers and children. In W. M. Reynolds & G. E. Miller (Eds.), *Handbook of child psychology: Educational psychology* (Vol. 7). Hoboken, NJ: Wiley.
- Pluye, P., & Hong, Q. N. (2014). Combining the power of stories and the power of numbers: Mixed methods research and mixed studies reviews. *Annual review of public health*, 35, 29-45. doi:10.1146/annurev-publhealth-032013-182440
- Polanin, J. R. (2013). *Addressing the issue of meta-analysis multiplicity in education and psychology*. (PhD), Loyola University Chicago. Retrieved from http://ecommons.luc.edu/luc_diss/539 (539)

- Reichardt, C. S., & Cook, T. D. (1979). Beyond qualitative versus quantitative methods. In T. D. Cook & C. S. Reichardt (Eds.), *Qualitative and quantitative methods in evaluation research* (Vol. 1, pp. 7-32). Beverly Hills, CA: SAGE.
- Reinke, W. M., Herman, K. C., & Stormont, M. (2010). Evaluation of a video-based modeling program to promote effective teacher classroom management practices: Institute of Education Sciences.
- Reinke, W. M., Herman, K. C., Stormont, M., Newcomer, L., & David, K. (2013). Illustrating the multiple facets and levels of fidelity of implementation to a teacher classroom management intervention. *Administration and Policy in Mental Health AND Mental Health Services Research*, 40, 494-506. doi:10.1007/s10488-013-0496-2
- Reinke, W. M., Stormont, M., Herman, K. C., & Newcomer, L. (2014). Using coaching to support teacher implementation of classroom-based interventions. *Journal of Behavioral Education*, 23, 150-167. doi:10.1007/s10864-013-9186-0
- Reinke, W. M., Stormont, M., Herman, K. C., Wang, Z., Newcomer, L., & King, K. (2014). Use of coaching and behavior support planning for students with disruptive behavior within a universal classroom management program. *Journal of Emotional and Behavioral Disorders*, 22, 74-82. doi:10.1177/1063426613519820
- Reinke, W. M., Stormont, M., Webster-Stratton, C., Newcomer, L. L., & Herman, K. C. (2012). The Incredible Years Teacher Classroom Management program: Using coaching to support generalization to real-world classroom settings. *Psychology in the Schools*, 49, 416-428. doi:10.1002/pits.21608
- Rockhill, K. (1983). Mandatory continuing education for professionals: Trends and issues. *Adult Education Quarterly*, 33, 106-116. doi:10.1177/0001848183033002003

- Rolfe, G. (2006). Validity, trustworthiness and rigour: Quality and the idea of qualitative research. *Journal of Advanced Nursing*, 53, 304-310. doi:10.1111/j.1365-2648.2006.03727.x
- Ross, S. W., Romer, N., & Horner, R. H. (2011). Teacher well-being and the implementation of schoolwide positive behavior interventions and supports. *Journal of Positive Behavior Interventions*, 14, 1098300711413820. doi:10.1177/1098300711413820
- Rossman, G. B., & Wilson, B. L. (1985). Numbers and words: Combining quantitative and qualitative methods in a single large-scale evaluation study. *Evaluation Review*, 9, 627-643. doi:10.1177/0193841X8500900505
- Russell, G., Golding, J., Norwich, B., Emond, A., Ford, T., & Steer, C. (2012). Social and behavioural outcomes in children diagnosed with autism spectrum disorders: A longitudinal cohort study. *Journal of child psychology and psychiatry*, 53, 735-744. doi:10.1111/j.1469-7610.2011.02490.x
- Sackett, D. L., Rosenberg, W. M., Gray, J. M., Haynes, R. B., & Richardson, W. S. (1996). Evidence based medicine: what it is and what it isn't. *BMJ*, 312, 71-72. doi:10.1136/bmj.312.7023.71
- Sammons, P. (2010). The contribution of mixed methods to recent research on educational effectiveness. In A. Tashakkori & C. Teddlie (Eds.), *SAGE Handbook of Mixed Methods in Social and Behavioral Research* (2nd ed., pp. 697-723). Thousand Oaks, CA: SAGE.
- Sandelowski, M., & Barroso, J. (2007). *Handbook for synthesizing qualitative research*. New York, NY: Springer Publishing Company.
- Schanzenbach, D. W. (2014). *Does class size matter*. Retrieved from Boulder, CO: <http://nepc.colorado.edu/publication/does-class-size-matter>

- Schulz, K. F., Altman, D. G., & Moher, D. (2010). CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *BMC medicine*, 8(1), 18.
- Shepard, S. A., Doctoroff, G. L., Dickstein, S., & Seifer, R. (2008). *An independent replication of the Incredible Years Classroom Management Teacher Training Program in Head Start*. Paper presented at the Society for Prevention Research.
- Snyder, H. (2001). Epidemiology of official offending. In R. Loeber & D. P. Farrington (Eds.), *Child delinquents: Development, intervention and service needs* (pp. 25-46). Thousand Oaks, CA: SAGE.
- Sonuga - Barke, E. J. (1998). Categorical models of childhood disorder: A conceptual and empirical analysis. *Journal of child psychology and psychiatry*, 39, 115-133.
doi:0021-9630/98
- Steckler, A. B., Linnan, L., & Israel, B. (2002). *Process evaluation for public health interventions and research*: Jossey-Bass San Francisco.
- Sturges, J. E., & Hanrahan, K. J. (2004). Comparing telephone and face-to-face qualitative interviewing: a research note. *Qualitative Research*, 4, 107-118.
doi:10.1177/1468794104041110
- Sugai, G., & Horner, R. R. (2006). A promising approach for expanding and sustaining school-wide positive behavior support. *School Psychology Review*, 35, 245.
- Tapp, J., Wehby, J., & Ellis, D. (1995). A multiple option observation system for experimental studies: MOOSES. *Behavior Research Methods, Instruments, & Computers*, 27, 25-31. doi:10.3758/BF03203616
- Teddle, C., & Sammons, P. (2010). Applications of mixed methods to the field of educational effectiveness research. In B. P. M. Creemers, L. Kyriakides, & P.

- Sammons (Eds.), *Methodological advances in educational effectiveness research*. London, UK: Routledge Taylor & Francis.
- Teddlie, C., & Tashakkori, A. (2009). *The foundations of mixed methods research: integrating quantitative and qualitative techniques in the social and behavioral sciences*. Thousand Oaks, CA: SAGE.
- The Incredible Years. (2013). Retrieved from <http://incredibleyears.com/programs/teacher/classroom-mgt-curriculum/>
- Tong, A., Flemming, K., McInnes, E., Oliver, S., & Craig, J. (2012). Enhancing transparency in reporting the synthesis of qualitative research: ENTREQ. *BMC Medical Research Methodology*, 12, 181-188. doi:10.1186/1471-2288-12-181
- Tong, A., Sainsbury, P., & Craig, J. (2007). Consolidated criteria for reporting qualitative research (COREQ): A 32-item checklist for interviews and focus groups. *International Journal for Quality in Health Care*, 19, 349-357. doi:10.1093/intqhc/mzm042
- UNESCO. (1994). *Salamanca statement and framework for action on special needs education*. Retrieved from
- Valentine, J. C., Pigott, T. D., & Rothstein, H. R. (2010). How many studies do you need? A primer on statistical power for meta-analysis. *Journal of Educational and Behavioral Statistics*, 35, 215-247. doi:10.3102/1076998609346961
- Van den Noortgate, W., López-López, J. A., Marín-Martínez, F., & Sánchez-Meca, J. (2013). Three-level meta-analysis of dependent effect sizes. *Behavior Research Methods*, 45, 576-594. doi:10.3758/s13428-012-0261-6
- Van Den Noortgate, W., & Onghena, P. (2003). Multilevel meta-analysis: A comparison with traditional meta-analytical procedures. *Educational and Psychological Measurement*, 63, 765-790. doi:10.1177/0013164403251027

- Veenman, S. (1984). Perceived problems of beginning teachers. *Review of educational research*, 54, 143-178. doi:10.3102/00346543054002143
- Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, 36, 1-48.
- Walsh, D., & Downe, S. (2005). Meta - synthesis method for qualitative research: a literature review. *Journal of Advanced Nursing*, 50(2), 204-211.
- Webster-Stratton, C. (2001). The Incredible Years: Parents, teachers, and children training series. *Residential Treatment for Children & Youth*, 18, 31-45.
doi:10.1300/J007v18n03_04
- Webster-Stratton, C. (2005). *Teacher Classroom Management Strategies Questionnaire*.
Retrieved from
http://www.google.co.uk/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=0ahUKEwiSgceEu9rNAhVYF8AKHQhuBf4QFghAMAA&url=http%3A%2F%2Fincredibleyears.com%2Fdownload%2Fresources%2Fteacher-pgrm%2FAmerican%2520Teacher%2520Strategies%2520Questionnaire.pdf&usq=AFQjCNFQ2ReAB_JWZ9jJrWAR_jEwfOqwSQ
- Webster-Stratton, C., Reid, M. J., & Hammond, M. (2001). Preventing conduct problems, promoting social competence: A parent and teacher training partnership in Head Start. *Journal of Clinical Child Psychology*, 30, 283-302.
doi:10.1207/S15374424JCCP3003_2
- Webster-Stratton, C., Reid, M. J., & Hammond, M. (2004). Treating children with early-onset conduct problems: Intervention outcomes for parent, child, and teacher training. *Journal of Clinical Child and Adolescent Psychology*, 33, 105-124.
doi:10.1207/S15374424JCCP3301_11

- Wetherall, M. (2014). Incredible Years Parent and Teacher Programmes: Emerging Themes and Issues. *Kairaranga*, 15, 36-41.
- Whear, R., Thompson-Coon, J., Boddy, K., Ford, T., Racey, D., & Stein, K. (2012). The effect of teacher-led interventions on social and emotional behaviour in primary school children: A systematic review. *British Educational Research Journal*, 39, 383-420. doi:10.1080/01411926.2011.650680
- Wickman, E. K. (1928). *Children's behavior and teachers' attitudes*. New York, NY: The Commonwealth Fund.
- Winter, E. C. (2006). Preparing new teachers for inclusive schools and classrooms. *Support for Learning*, 21, 85-91. doi:10.1111/j.1467-9604.2006.00409.x
- Wittchen, H.-U., Jacobi, F., Rehm, J., Gustavsson, A., Svensson, M., Jönsson, B., . . . Faravelli, C. (2011). The size and burden of mental disorders and other disorders of the brain in Europe 2010. *European Neuropsychopharmacology*, 21, 655-679. doi:10.1016/j.euroneuro.2011.07.018
- World Health Organization. (1992). *The ICD-10 classification of mental and behavioural disorders: clinical descriptions and diagnostic guidelines*. Geneva: World Health Organization.

Appendix A: PROSPERO Protocol for Study One

PROSPERO International prospective register of systematic reviews

Mixed methods systematic review of a teacher classroom management programme: effectiveness and stakeholders' experiences

Elizabeth Nye, G. J. Melendez-Torres, Frances Gardner

Citation

Elizabeth Nye, G. J. Melendez-Torres, Frances Gardner. Mixed methods systematic review of a teacher classroom management programme: effectiveness and stakeholders' experiences. PROSPERO 2015:CRD42015020359
Available from http://www.crd.york.ac.uk/PROSPERO_REBRANDING/display_record.asp?ID=CRD42015020359

Nye, E., Melendez-Torres, G. J., Gardner, F. (2015). Mixed methods systematic review of a teacher classroom management programme: effectiveness and stakeholders' experiences. PROSPERO 2015: CRD42015020359. Available from http://www.crd.york.ac.uk/PROSPERO/display_record.asp?ID=CRD42015020359.

Review title and timescale

1 Review title

Give the working title of the review. This must be in English. Ideally it should state succinctly the interventions or exposures being reviewed and the associated health or social problem being addressed in the review.

Mixed methods systematic review of a teacher classroom management programme: effectiveness and stakeholders' experiences

2 Original language title

For reviews in languages other than English, this field should be used to enter the title in the language of the review. This will be displayed together with the English language title.

3 Anticipated or actual start date

Give the date when the systematic review commenced, or is expected to commence.

01/07/2015

4 Anticipated completion date

Give the date by which the review is expected to be completed.

30/06/2016

5 Stage of review at time of this submission

Indicate the stage of progress of the review by ticking the relevant boxes. Reviews that have progressed beyond the point of completing data extraction at the time of initial registration are not eligible for inclusion in PROSPERO. This field should be updated when any amendments are made to a published record.

The review has not yet started ☒

Review stage	Started	Completed
Preliminary searches	Yes	Yes
Piloting of the study selection process	Yes	No
Formal screening of search results against eligibility criteria	Yes	No
Data extraction	No	No
Risk of bias (quality) assessment	No	No
Data analysis	No	No

Provide any other relevant information about the stage of the review here.

Review team details

6 Named contact

The named contact acts as the guarantor for the accuracy of the information presented in the register record.

Ms Nye

7 Named contact email

Enter the electronic mail address of the named contact.

elizabeth.nye@spi.ox.ac.uk

8 Named contact address

Enter the full postal address for the named contact.

Centre for Evidence-Based Intervention, Barnett House, 32 Wellington Square,
Oxford, OX1 2ER, United Kingdom

9 Named contact phone number

Enter the telephone number for the named contact, including international dialing code.

10 Organisational affiliation of the review

Full title of the organisational affiliations for this review, and website address if available. This field may be completed as 'None' if the review is not affiliated to any organisation.

Centre for Evidence-Based Intervention, University of Oxford

Website address:

<http://www.cebi.ox.ac.uk/home.html>

11 Review team members and their organisational affiliations

Give the title, first name and last name of all members of the team working directly on the review. Give the organisational affiliations of each member of the review team.

Title	First name	Last name	Affiliation
Ms	Elizabeth	Nye	University of Oxford
Dr	G. J.	Melendez-Torres	University of Warwick
Professor	Frances	Gardner	University of Oxford

12 Funding sources/sponsors

Give details of the individuals, organizations, groups or other legal entities who take responsibility for initiating, managing, sponsoring and/or financing the review. Any unique identification numbers assigned to the review by the individuals or bodies listed should be included.

None

13 Conflicts of interest

List any conditions that could lead to actual or perceived undue influence on judgements concerning the main topic investigated in the review.

Are there any actual or potential conflicts of interest?

None known

14 Collaborators

Give the name, affiliation and role of any individuals or organisations who are working on the review but who are not listed as review team members.

Title	First name	Last name	Organisation details
-------	------------	-----------	----------------------

Review methods

15 Review question(s)

State the question(s) to be addressed / review objectives. Please complete a separate box for each question.

What are the effects on teachers' management strategies and children's behaviour of the Incredible Years Teacher Classroom Management Programme compared to treatment-as-usual, at post-intervention and any measured follow-up time points?

What are the experiences and views of key stakeholders (e.g., teachers, school leaders, parents, children) of the Incredible Years Teacher Classroom Management Programme (e.g., content, processes, effects)?

When considered in combination, what can the effectiveness and stakeholders' experience data tell us?

16 Searches

Give details of the sources to be searched, and any restrictions (e.g. language or publication period). The full search strategy is not required, but may be supplied as a link or attachment.

Using a comprehensive search strategy for the intervention terms applied to full/all text, we will search the following electronic databases: 1. Applied Social Sciences Index and Abstracts (ASSIA) 2. British Education Index 3. Cochrane Central Register of Controlled Trials (CENTRAL) 4. CINAHL 5. EMBASE 6. ERIC 7. Linguistics and Language Behaviour Abstracts 8. MEDLINE 9. PAIS International 10. ProQuest Dissertations and Theses 11. Social Services Abstracts 12. Sociological Abstracts 13. PsycINFO We will also search the following website article libraries: 1. Campbell Collaboration Library 2. Cochrane Collaboration Library 3. Incredible Years Series (<http://incredibleyears.com/>) 4. Centre for Evidence-Based Early Intervention (Wales) The search will include Open Grey and emails to researchers in the field. Searches will be conducted in English, but we will apply no language restrictions.

17 URL to search strategy

If you have one, give the link to your search strategy here. Alternatively you can e-mail this to PROSPERO and we will store and link to it.

I give permission for this file to be made publicly available

Yes

18 Condition or domain being studied

Give a short description of the disease, condition or healthcare domain being studied. This could include health and wellbeing outcomes.

This review examines a teacher classroom management programme (Incredible Years)

that addresses child behavioural, social-emotional, and academic difficulties (e.g., conduct disorder, mental health, peer relationships) and teacher's practices (e.g., positive and/or negative behaviour management strategies). Conduct disorder among children and adolescents has been established at 5.8 per cent in the UK (Green et al., 2005), 3 per cent in a broad review of European studies (Wittchen et al., 2011), and 2.1 per cent in the US (Merikangas et al., 2010). Children identified with problematic behaviours (not reaching clinical levels of conduct disorder) add another roughly 20 per cent (Murray et al., 2013), meaning nearly a quarter of children demonstrate difficulties in this area. These difficulties present challenges for educators, as children's problematic behaviours have been correlated with reduced academic performance, stressed pupil-teacher relationships, poorer peer interactions, and increased chances of dropping out of school before completion (Baker et al., 2009; McMahon et al., 2006).

19 Participants/population

Give summary criteria for the participants or populations being studied by the review.

The preferred format includes details of both inclusion and exclusion criteria.

The Incredible Years Teacher Classroom Management (IY TCM) Programme is designed for early childhood and primary school teachers of children ages three to eight years old. Included quantitative studies will have implemented IY TCM with this population. Qualitative studies of the experiences of key stakeholders include people who have either direct contact with the IY TCM Programme (e.g., teachers, group leaders) or who have indirect contact via implementation of the programme's strategies (e.g., other educators, school children and their parents).

20 Intervention(s), exposure(s)

Give full and clear descriptions of the nature of the interventions or the exposures to be reviewed

The Webster-Stratton Incredible Years Series is a suite of training programmes aimed at parents, teachers, and children to decrease child aggression and conduct problems and increase parent/teacher competencies and child prosocial behaviours. The Incredible Years Teacher Classroom Management (IY TCM) Programme is one of the packages developed to address conduct disorder via training teachers to better address issues in the school environment. It has six components: 1. The importance of teacher attention, encouragement, and praise 2. Motivating children through incentives 3.

Preventing problems—The proactive teacher 4. Decreasing inappropriate behaviour 5. Building positive relationships with children 6. Social skills and problem solving training in the classroom (Webster-Stratton, 2001). Trained facilitators lead group sessions for teachers over six monthly full-day sessions. Video vignettes of teachers in full-size classrooms (28 children) and of special education classrooms with smaller teacher to pupil ratios introduce the topics for discussion. Teachers problem solve collaboratively in the group sessions before heading back into their classroom to implement the learned strategies (Webster-Stratton, 2001).

21 Comparator(s)/control

Where relevant, give details of the alternatives against which the main subject/topic of the review will be compared (e.g. another intervention or a non-exposed control group).

For the quantitative studies included in this review, IY TCM must be compared to a control group that offers either wait-list control or treatment-as-usual. There is no control group necessary for the qualitative studies included in this review.

22 Types of study to be included

Give details of the study designs to be included in the review. If there are no restrictions on the types of study design eligible for inclusion, this should be stated.

The quantitative strand of this review includes all relevant individual- or cluster-randomized controlled trials (RCTs). This review is limited to studies using an RCT experimental design because its focus is on evaluating intervention outcomes.

Randomly assigning participants to treatment and control groups minimizes potential biases by making the different groups equivalent on known and unknown variables that could otherwise influence intervention outcomes (Fraser et al., 2009). It is not uncommon in school-based interventions to use cluster-randomized controlled trials, with the unit of allocation potentially at the level of teacher/classroom, school, or other natural grouping, as this serves to minimize treatment contamination and because the intervention is delivered and designed for impact at the group level (Fraser et al., 2009). Studies should examine IY TCM in isolation from other programmes and studies will be excluded if IY TCM is not isolated (e.g., delivered with IY parent training or other school interventions that are not typically provided in the studies' contexts). The qualitative strand of this review includes evidence from either qualitative or mixed methods studies that report on experiences and views of

key stakeholders. Qualitative studies associated with RCTs will be included. Experiences or views that are reported in a quantitative manner will be excluded from the qualitative strand of this review.

23 Context

Give summary details of the setting and other relevant characteristics which help define the inclusion or exclusion criteria.

24 Primary outcome(s)

Give the most important outcomes.

Outcomes might be measured by independent observers, teacher-reports, parent-reports, or other sources including school records. Including results across these different types of measures allows for balancing the individual weaknesses inherent in each type alone. Primary outcomes: 1. Teacher: classroom management strategies (positive and negative) 2. Child: conduct problems (e.g., aggressive and noncompliant behaviors) The qualitative strand of this review will explore social processes and seek to understand the interactions of key stakeholders with the content and delivery of the IY TCM Programme and their subsequent views of the strategies and implementation of the techniques in their practice.

Give information on timing and effect measures, as appropriate.

This review measures immediate trial outcomes (post-intervention) and any follow-up outcomes as available. Qualitative findings might be based from data collected during or after IY TCM was implemented. Qualitative findings collected before implementation of IY TCM will only be included if they directly address the IY TCM Programme (as opposed to classroom management in general or other topics), because this review is interested in stakeholders' experiences with and views on IY TCM.

25 Secondary outcomes

List any additional outcomes that will be addressed. If there are no secondary outcomes enter None.

Secondary outcomes: 1. Teacher: collaboration with parents 2. Child: prosocial behaviors 3. Child: academic readiness (e.g., reading skills)

Give information on timing and effect measures, as appropriate.

This review measures immediate trial outcomes (post-intervention) and any follow-up outcomes as available.

26 Data extraction (selection and coding)

Give the procedure for selecting studies for the review and extracting data, including the number of researchers involved and how discrepancies will be resolved. List the data to be extracted.

Data will be extracted by the lead author and checked by the second. Discrepancies will be resolved in consultation with the third author. All teacher and child outcomes will be extracted from included trial studies. All stakeholder experiences and views of IY TCM will be extracted for the qualitative metasynthesis.

27 Risk of bias (quality) assessment

State whether and how risk of bias will be assessed, how the quality of individual studies will be assessed, and whether and how this will influence the planned synthesis.

Studies will be assessed for risk of bias (quantitative) and quality (qualitative) following Cochrane guidelines. Where necessary, we will request additional information from authors of included studies to make informed appraisal decisions. Quantitative studies found to be at high-risk for bias will be separated out in sensitivity analyses to determine any undue effect on overall outcomes. Qualitative studies will be assessed for quality using the EPPI-Centre checklist. Qualitative studies will not be excluded on the basis of a poor quality assessment; however, such studies will not serve as the initial point for reciprocal translation in the metasynthesis process.

28 Strategy for data synthesis

Give the planned general approach to be used, for example whether the data to be used will be aggregate or at the level of individual participants, and whether a quantitative or narrative (descriptive) synthesis is planned. Where appropriate a brief outline of analytic approach should be given.

Both quantitative and qualitative included studies will be narratively described to start. Additionally, a quantitative meta-analysis will examine the overall effects of the trial data. As with many school-based interventions there is a natural clustering effect, whereby the organisation of the educational system creates a hierarchy of levels within which interventions are delivered (Goldstein, 1987). Operating from the logical assumption that such a hierarchical structure precludes independence at any given level, the meta-analysis of quantitative trials will follow a multilevel model. A

multilevel model estimates residual variance of outcomes at a lower level in relation to both the effect of higher levels on the lower level outcomes as well as to any unexplained error (Bateman & Jones, 2003). This is important for the results of meta-analyses, because large intracluster correlations subsequently result in over-rejection of the null-hypothesis for under-estimating standard errors (Goldstein, 1987; Bateman & Jones, 2003). An interpretive qualitative metasynthesis based in grounded theory approach will serve as further analysis of included qualitative studies. Through coding, qualitative data is transformed from its original state to a conceptual level (Corbin & Strauss, 2008). To start, open codes are assigned to the data, with the goal of identifying the many various inquiries one could pursue with the data (Green & Thorogood, 2014). Then, categories are connected by establishing their relationships to one another through axial coding. The analysis proceeds towards progressively more abstract and theoretically informed codes with the emergence of core categories that remain substantiated in the original data.

29 Analysis of subgroups or subsets

Give any planned exploration of subgroups or subsets within the review. 'None planned' is a valid response if no subgroup analyses are planned.

Subgroup analyses are planned for children deemed to be 'at-risk' with high baseline levels of problematic behaviours. Themes arising from the metasynthesis will be considered in relation to the various stakeholders whose experiences/views contributed to the findings.

Review general information

30 Type and method of review

Select the type of review and the review method from the drop down list.

Intervention, Systematic review, Other

31 Language

Select the language(s) in which the review is being written and will be made available, from the drop down list. Use the control key to select more than one language.

English

Will a summary/abstract be made available in English?

Yes

32 Country

Select the country in which the review is being carried out from the drop down list.
For multi-national collaborations select all the countries involved. Use the control key to select more than one country.

England

33 Other registration details

Give the name of any organisation where the systematic review title or protocol is registered together with any unique identification number assigned. If extracted data will be stored and made available through a repository such as the Systematic Review Data Repository (SRDR), details and a link should be included here.

34 Reference and/or URL for published protocol

Give the citation for the published protocol, if there is one.

Give the link to the published protocol, if there is one. This may be to an external site or to a protocol deposited with CRD in pdf format.

I give permission for this file to be made publicly available

Yes

35 Dissemination plans

Give brief details of plans for communicating essential messages from the review to the appropriate audiences.

Do you intend to publish the review on completion?

Yes

36 Keywords

Give words or phrases that best describe the review. (One word per box, create a new box for each term)

mixed methods systematic review

multilevel meta-analysis

qualitative metasynthesis

teacher classroom management

37 Details of any existing review of the same topic by the same authors

Give details of earlier versions of the systematic review if an update of an existing review is being registered, including full bibliographic reference if possible.

38 Current review status

Review status should be updated when the review is completed and when it is published.

Ongoing

39 Any additional information

Provide any further information the review team consider relevant to the registration of the review.

40 Details of final report/publication(s)

This field should be left empty until details of the completed review are available.

Give the full citation for the final report or publication of the systematic review.

Give the URL where available.

Appendix B: R Syntax for multilevel meta-analysis for Study One

```
>####Setting the working directory###
>setwd("Z:/1DPhil/7MMSR Study/R_Working_Directory")
>
>####Teacher Negative Strategies###
>library(metafor)
>TNS<-read.csv("TNS.csv", header=T)
>TNS
>MLM_TNS<-rma.mv(yi=hedges_g, V=V_g, data=TNS, random=list(~1|esid, ~1|studyid),
slab=studyoutcome)
>summary(MLM_TNS)
>I2_TNS_es<-
(MLM_TNS$sigma2[1])/(MLM_TNS$sigma[1]+MLM_TNS$sigma[2]+mean(TNS$V_g))
>I2_TNS_study<-
(MLM_TNS$sigma2[2])/(MLM_TNS$sigma[1]+MLM_TNS$sigma[2]+mean(TNS$V_g))
>I2_TNS_es
>I2_TNS_study
>
>####Teacher Positive Strategies###
>library(metafor)
>TPS<-read.csv("TPS.csv", header=T)
>TPS
>MLM_TPS<-rma.mv(yi=hedges_g, V=V_g, data=TPS, random=list(~1|esid, ~1|studyid),
slab=studyoutcome)
>summary(MLM_TPS)
>I2_TPS_es<-
(MLM_TPS$sigma2[1])/(MLM_TPS$sigma[1]+MLM_TPS$sigma[2]+mean(TPS$V_g))
>I2_TPS_study<-
(MLM_TPS$sigma2[2])/(MLM_TPS$sigma[1]+MLM_TPS$sigma[2]+mean(TPS$V_g))
>I2_TPS_es
>I2_TPS_study
>
>####Conduct Problems###
>library(metafor)
>ConProb<-read.csv("ConProb.csv", header=T)
>ConProb
>MLM_ConProb<-rma.mv(yi=hedges_g, V=V_g, data=ConProb, random=list(~1|esid,
~1|studyid), slab=studyoutcome)
>summary(MLM_ConProb)
>I2_ConProb_es<-
(MLM_ConProb$sigma2[1])/(MLM_ConProb$sigma[1]+MLM_ConProb$sigma[2]+mean(ConPr
ob$V_g))
```

```

>I2_ConProb_study<-
(MLM_ConProb$sigma2[2])/((MLM_ConProb$sigma[1]+MLM_ConProb$sigma[2]+mean(ConProb$V_g))
>I2_ConProb_es
>I2_ConProb_study
>
>####Conduct Problems - Sensitivity####
>library(metafor)
>ConProb_Sens<-read.csv("ConProb_Sens.csv", header=T)
>ConProb_Sens
>MLM_ConProb_Sens<-rma.mv(yi=hedges_g, V=V_g, data=ConProb_Sens,
random=list(~1|esid, ~1|studyid), slab=studyoutcome)
>summary(MLM_ConProb_Sens)
>I2_ConProb_Sens_es<-
(MLM_ConProb_Sens$sigma2[1])/((MLM_ConProb_Sens$sigma[1]+MLM_ConProb_Sens$sigma
[2]+mean(ConProb_Sens$V_g))
>I2_ConProb_Sens_study<-
(MLM_ConProb_Sens$sigma2[2])/((MLM_ConProb_Sens$sigma[1]+MLM_ConProb_Sens$sigma
[2]+mean(ConProb_Sens$V_g))
>I2_ConProb_Sens_es
>I2_ConProb_Sens_study
>
>####Conduct Problems - High Risk####
>library(metafor)
>ConProb_High<-read.csv("ConProb_High.csv", header=T)
>ConProb_High
>MLM_ConProb_High<-rma.mv(yi=hedges_g, V=V_g, data=ConProb_High,
random=list(~1|esid, ~1|studyid), slab=studyoutcome)
>summary(MLM_ConProb_High)
>I2_ConProb_High_es<-
(MLM_ConProb_High$sigma2[1])/((MLM_ConProb_High$sigma[1]+MLM_ConProb_High$sigma
[2]+mean(ConProb_High$V_g))
>I2_ConProb_High_study<-
(MLM_ConProb_High$sigma2[2])/((MLM_ConProb_High$sigma[1]+MLM_ConProb_High$sigma
[2]+mean(ConProb_High$V_g))
>I2_ConProb_High_es
>I2_ConProb_High_study
>
>####Prosocial Behaviour####
>library(metafor)
>Prosoc<-read.csv("Prosoc.csv", header=T)
>Prosoc

```

```

>MLM_Prococ<-rma.mv(yi=hedges_g, V=V_g, data=Prosoc, random=list(~1|esid,
~1|studyid), slab=studyoutcome)
summary(MLM_Prococ)
>I2_Prococ_es<-
(MLM_Prococ$sigma2[1])/(MLM_Prococ$sigma[1]+MLM_Prococ$sigma[2]+mean(Prosoc$V_g))
>I2_Prococ_study<-
(MLM_Prococ$sigma2[2])/(MLM_Prococ$sigma[1]+MLM_Prococ$sigma[2]+mean(Prosoc$V_g))
>I2_Prococ_es
>I2_Prococ_study
>
>####Prosocial Behaviour - High Risk###
>library(metafor)
>Prosoc_High<-read.csv("Prosoc_High.csv", header=T)
>Prosoc_High
>MLM_Prococ_High<-rma.mv(yi=hedges_g, V=V_g, data=Prosoc_High,
random=list(~1|esid, ~1|studyid), slab=studyoutcome)
>summary(MLM_Prococ_High)
>I2_Prococ_High_es<-
(MLM_Prococ_High$sigma2[1])/(MLM_Prococ_High$sigma[1]+MLM_Prococ_High$sigma[2]+mean(Prosoc_High$V_g))
>I2_Prococ_High_study<-
(MLM_Prococ_High$sigma2[2])/(MLM_Prococ_High$sigma[1]+MLM_Prococ_High$sigma[2]+mean(Prosoc_High$V_g))
>I2_Prococ_High_es
>I2_Prococ_High_study
>####Behaviour Difficulties - Sensitivity###
>library(metafor)
>BehDiff_Sens<-read.csv("BehDiff_Sens.csv", header=T)
>BehDiff_Sens
>MLM_BehDiff_Sens<-rma.mv(yi=hedges_g, V=V_g, data=BehDiff_Sens,
random=list(~1|esid, ~1|studyid), slab=studyoutcome)
>summary(MLM_BehDiff_Sens)
>I2_BehDiff_Sens_es<-
(MLM_BehDiff_Sens$sigma2[1])/(MLM_BehDiff_Sens$sigma[1]+MLM_BehDiff_Sens$sigma[2]+mean(BehDiff_Sens$V_g))
>I2_BehDiff_Sens_study<-
(MLM_BehDiff_Sens$sigma2[2])/(MLM_BehDiff_Sens$sigma[1]+MLM_BehDiff_Sens$sigma[2]+mean(BehDiff_Sens$V_g))

```

```

>l2_BehDiff_Sens_es
>l2_BehDiff_Sens_study
>
>###Behaviour Difficulties - High Risk###
>library(metafor)
>BehDiff_High<-read.csv("BehDiff_High.csv", header=T)
>BehDiff_High
>MLM_BehDiff_High<-rma.mv(yi=hedges_g, V=V_g, data=BehDiff_High,
random=list(~1|esid, ~1|studyid), slab=studyoutcome)
>summary(MLM_BehDiff_High)
>l2_BehDiff_High_es<-
(MLM_BehDiff_High$sigma2[1])/(MLM_BehDiff_High$sigma[1]+MLM_BehDiff_High$si
gma[2]+mean(BehDiff_High$V_g))
>l2_BehDiff_High_study<-
(MLM_BehDiff_High$sigma2[2])/(MLM_BehDiff_High$sigma[1]+MLM_BehDiff_High$si
gma[2]+mean(BehDiff_High$V_g))
>l2_BehDiff_High_es
>l2_BehDiff_High_study

```


Appendix C: First-authored peer-reviewed journal article on qualitative meta-synthesis associated with Study One

Review of Education
An International Journal of Major Studies in Education

BERA
BRITISH EDUCATIONAL RESEARCH ASSOCIATION

[Explore this journal >](#)

Original Article

Origins, methods and advances in qualitative meta-synthesis

Elizabeth Nye [✉](#), G. J. Melendez-Torres, Chris Bonell

First published: February 2016 [Full publication history](#)

DOI: 10.1002/rev3.3065 [View/save citation](#)

This is the accepted version of the following article: Nye, E., Melendez-Torres, G. J., & Bonell, C. (2016). Origins, methods and advances in qualitative meta-synthesis. Review of Education, 4(1), 57-79, which has been published in final form at [http://onlinelibrary.wiley.com/doi/10.1002/rev3.3065/full]. This article may be used for non-commercial purposes in accordance with the Wiley Self-Archiving Policy [http://olabout.wiley.com/WileyCDA/Section/id-820227.html].

Origins, methods and advances in qualitative meta-synthesis

Elizabeth Nye^{1*}, G. J. Melendez-Torres², and Chris Bonell³

¹*Department of Social Policy and Intervention, University of Oxford, Oxford, UK*

²*Warwick Evidence, Division of Health Sciences, Warwick Medical School, University of Warwick, Coventry, UK*

³*Social Science Research Unit, Department of Social Science, UCL Institute of Education, London, UK*

**Correspondence to Elizabeth Nye, Barnett House, 32 Wellington Square, Oxford, OX1 2ER, UK. Email: elizabeth.nye@spi.ox.ac.uk.*

Abstract

Qualitative research is a broad term encompassing many methods (Dixon-Woods, Fitzpatrick, & Roberts, 2001; Evans, 2002). Critiques of the field of qualitative research argue that while individual studies provide rich descriptions and insights, the absence of connections drawn between studies limits their usefulness. In response, qualitative meta-

synthesis serves as a design to interpret and synthesise qualitative findings across individual studies. More than a broad summary, meta-syntheses do not aim merely to summarise all available data; rather, qualitative meta-syntheses present new perspectives on topics through interpreting findings from different qualitative studies to create ‘third-level’ findings for the advancement of both knowledge (Jensen & Allen, 1996; Sandelowski, 1993) and theory (Atkins et al., 2008; Glaser & Strauss, 1967). The diversity of opinion on qualitative meta-synthesis is mirrored in its practice. Several different approaches to qualitative meta-synthesis have emerged, with most connected to the meta-ethnographic procedures originally outlined by educationalists Noblit and Hare (1988). This paper: (1) discusses the key philosophical and methodological issues in the literature on qualitative meta-synthesis, (2) highlights key methods that are used in qualitative meta-synthesis, and (3) offers an overview of where the field is going. Recent examples of qualitative meta-syntheses highlight some of this design’s current contributions and future usefulness for research in the field of education.

Keywords: qualitative meta-synthesis, meta-ethnography, reciprocal translation, lines-of-argument synthesis

Introduction

This paper addresses qualitative meta-synthesis as it has been applied in the field of education. Broadly speaking, qualitative meta-synthesis is the systematic review and additional level of interpretation of primary qualitative research studies. While much of the recent methodological development of meta-synthesis has occurred in the field of public health, this line of research originates with educationalists Noblit and Hare (1988) and their germinal book on meta-ethnography. Meta-ethnography—or, the translation of concepts and metaphors across studies—has thus formed the basis for subsequent iterations and adaptations of qualitative meta-synthesis, and we will examine meta-ethnography in the context of the historical impetus for qualitative meta-synthesis. Then, we will discuss the key philosophical and methodological issues in the literature on qualitative meta-synthesis, highlight key methods that are used in qualitative meta-synthesis, and offer an overview of where the field is going. Throughout this paper we highlight recent examples of qualitative meta-syntheses in education to demonstrate some of this design’s current contributions and its usefulness for future research in the field of education.

A brief description of qualitative methodologies

Entrenched in an interpretive paradigm that perceives multiple, socially constructed realities (Jensen & Allen, 1996; Sandelowski, 1993), qualitative research is the line of inquiry into how people interact with and interpret the world (Atkins et al., 2008; Barroso et al., 2003). Giddens (1984) describes this as a double hermeneutic, which highlights the multiple layers and bidirectionality of people interpreting their everyday world and social scientists' interpretations of those interpretations. Indeed, qualitative research is 'the exploration of meanings of social phenomena as experienced by individuals themselves, in their natural context' (Malterud, 2001, p. 483). Qualitative research is a broad term encompassing a variety of approaches, such as ethnography, and methods, including focus groups, interviewing, and observation (Dixon-Woods et al., 2001; Evans, 2002). As a whole, qualitative methods aim to understand meaning rather than quantify the effects or distributions of parameters (Green & Thorogood, 2009). Therefore, qualitative research deals with non-numerical data analysis, focusing instead on narrative data analysis through themes and concepts (Dixon-Woods et al., 2001; Evans, 2002).

Qualitative research is hardly monolithic in its intersecting and overlapping approaches to data collection and analysis. Overall, a common thread is an interpretive epistemology—a theory of knowledge that aims to understand, for example, how people interpret a phenomenon, perceive a particular issue or problem, or account for a certain behaviour (Green & Thorogood, 2009). This interpretive epistemology has been called 'hermeneutic', or understanding the object of study as both parts and whole through experiences and understandings contained in collected data (Zimmer, 2006). For example, social realists believe that collected data represents a concrete, commonly shared reality—a perspective arguably more positivistic than interpretive (Fine, 1993). Critical realism modifies social realism to note that phenomena exist at different 'levels', including the level at which they concretely occur and the level at which people experience them, with qualitative data focusing on accounts of these phenomena (Houston, 2001). Social constructionism dispenses with the idea of commonly experienced reality and instead studies how people construct this reality (Green & Thorogood, 2009; Riegler, 2005). A 'strong' version of social constructionism holds that all in the world is socially constructed, while a 'weak' version of this epistemology addresses the construction of social structures and phenomena (Houston, 2001). Phenomenology addresses the specific 'lifeworlds' of interview subjects by understanding how they relate to and understand

specific phenomena, often through presentation of substantial detail and examination of how subjects process their own understandings (Schutz, 1962; Suddaby, 2006). Symbolic interactionism examines how the ‘situated self’ understands the different dimensions of the lived world as a set of interactions through which meaning is produced and ascribed, with these ascribed meanings called ‘symbols’ (Fine, 1993). These meanings produce interactions, themselves altering the meanings informing them (Blumer, 1992).

Grounded theory—one of the most widely used approaches to qualitative data analysis—was developed by the symbolic interactionists Barney Glaser and Anselm Strauss. Research oriented by a symbolic interactionist framework takes as the object of analysis the meanings embedded in the interactions between actors, and how these intersect to produce a social process (Corbin & Strauss, 2008). Because symbolic interactionism is a framework rather than an epistemology, researchers have used various interpretive epistemologies in grounded theory research. While Glaser (2002) posits that grounded theory can be used across epistemologies, Charmaz (2006) argues that grounded theory must be oriented in a social constructionist epistemology to remain true to its original conception. On the other hand, several major grounded theory exponents, while preserving a symbolic interactionist framework, have moved towards an essentially positivist approach (Annells, 1997; Corbin & Strauss, 2008).

Grounded theory analysis begins with open coding, a close reading of small parts of the data to identify common concepts in different sources, and to understand data in depth (Green & Thorogood, 2009). From this, a starting set of codes—or labels attached to the data that identify an idea the data expresses—is systematically assigned to all the data fragment-by-fragment. This coding scheme is continuously updated through constant comparison with newly analysed data during ‘axial coding’, in which the data are examined to see how codes relate to each other and point to underlying, higher-order ‘concepts’, into which codes are collapsed (Corbin & Strauss, 2008). Corbin and Strauss then describe the process of generating ‘categories’, an even higher-order set of abstractions that are theoretically saturated collections of concepts. These categories have ‘properties’—concepts representing different aspects of the category—and ‘dimensions’, concepts representing how the properties of the category might vary across different respondents. Throughout analysis, the process should be methodologically auditable, both through generation of frequent memos recorded by the analyst and through records of decisions from open coding to the final theory (Corbin & Strauss, 2008; Finfgeld-Connett,

2010; Kearney, 2001). Introduced in the 1960s, grounded theory became a dominant force in Western qualitative research (Thorne, Jensen, Kearney, Noblit, & Sandelowski, 2004).

Three other approaches to qualitative data analysis deserve brief mention here. The first is thematic analysis, which is a method for the identification of themes within the data. This method proceeds systematically through interviewing; examination of data; clustering into themes that are relevant, auditable, and reliable; and presentation of these themes (Saldana, 2009). The second method, which is closely related to thematic analysis, is framework analysis, originally developed for rapid analysis and policy relevance (Green & Thorogood, 2009). Framework analysis proceeds similarly, but accommodates large amounts of data through mapping of identified themes. Themes may be defined inductively or deductively, depending on the research team's needs (Pope, Ziebland, & Mays, 2000). Finally, Miles and Huberman present a method of cross-case analysis that proceeds through the steps of 'data reduction', 'data display', and formulation of conclusions (Miles & Huberman, 1984, 1994). This method is helpful for comparing and contrasting data both within and across cases. While all three methods are relevant for elaboration of recurring patterns in the data, they do not invite consideration of underlying themes, perceptions, and processes in the way grounded theory does. The inductive and deductive processes of searching for latent concepts and categories through open coding and axial coding, the imperative to theoretical saturation, and the symbolic interactionist framework orienting grounded theory towards an explicit examination of social structures invites a depth and focus of analysis other methods might lack. Yet in practice a great deal of qualitative research described as informed by grounded theory falls short of this standard.

Across approaches, qualitative research focuses on descriptive, rather than statistical, power. Qualitative studies often eschew large sample sizes for samples capturing a diversity of thought and approaches. Studies frequently stop data collection when data no longer generates substantively new insights (Bowling, 2009). Glaser and Strauss (1967) labelled this aspect of qualitative research 'theoretical saturation'. Qualitative researchers often seek to capture a broad range of experiences in an effort to find the unifying themes and concepts that tie a variety of experiences together, or that separate some participants from others. Indeed, this is a key distinction between qualitative research and quantitative research, which seeks a sample size that is both large enough to test hypotheses and representative enough to have a defined external validity.

Instead, qualitative research aims for ‘analytic generalisability’, wherein the study results are analysed in light their conceptual and descriptive power, and whether they describe hypotheses consistent with other studies (Mills, Durepos, & Wiebe, 2010). This prioritisation of conceptual richness over statistical generalisability poses a problem for those seeking to comprehend and simultaneously understand the findings of several qualitative studies that point to the same object of study.

Foundations and logic of qualitative meta-synthesis

The field of qualitative research currently lacks a collective outlook spanning various qualitative findings, despite an increasing amount of published studies (Britten et al., 2002). Critiques directed towards the field of qualitative research argue that while individual studies provide rich descriptions and insights of their selective group of focus, the absence of connections drawn between studies limits their usefulness in understanding a phenomenon (Jensen & Allen, 1996), informing practice (Zimmer, 2006), advising policy (Harden et al., 2004), and developing theory (Estabrooks, Field, & Morse, 1994). Indeed the current situation of qualitative studies operating as stand-alone pieces of evidence has led to the terms ‘cottage industry’ (Sandelowski, Docherty, & Emden, 1997, p. 366), ‘one-shot research’ (Estabrooks et al., 1994, p. 510), and ‘little islands of knowledge’ (Glaser & Strauss, 1971, p. 181).

In response to these criticisms, qualitative meta-synthesis serves as a method of interpreting and synthesising qualitative findings across individual studies. More than simply a broad summary of previous findings, meta-syntheses of qualitative research are not meant to ‘sum’ all available data; rather, meta-syntheses present new perspectives on topics through interpreting findings from different qualitative studies to create ‘third-level’ findings for the advancement of both knowledge (Jensen & Allen, 1996; Sandelowski, 1993) and theory (Atkins et al., 2008; Glaser & Strauss, 1967).

A commonly cited rationale (Kearney, 2001; Sandelowski et al., 1997; Walsh & Downe, 2005) for qualitative meta-synthesis arises from ‘analytic interruptus’. These meta-synthesists find ‘interruptus’ in the lack of horizontal relationships between qualitative studies to aid in transferring concepts between different studies’ accounts. However, Lofland (1970) initially coined ‘analytic interruptus’ not to describe relationships between study-level understandings, but how qualitative research does not carry through to broader ‘middle-range’ theories. Lofland described qualitative research as ‘conceptually impoverished’, highlighting the vast gap in theoretical range between

large-scale grand theories and the small, context-specific descriptions popular in accounts of qualitative research. The ‘interruptus’, as discussed by Lofland, occurs in the failure to ‘follow through to the implied, logical, or entailed conclusion’, the higher-order understandings that characterise ‘middle-range’ theory. Elsewhere, Strauss (1970) describes the process of ‘discovering new theory from previous theory’, or the building of theory from initial accounts by adding data and developing increasingly nuanced understandings. Strauss was more concerned with horizontal links between studies, and between accounts within studies. Yet his comment applies to qualitative meta-synthesis, in that the existing theory developed by individual studies can be used to develop new theories.

Qualitative meta-synthesis first came to prominence with educationalists Noblit and Hare’s (1988) *Meta-Ethnography: Synthesizing Qualitative Studies*. Though Noblit and Hare were not the first to write on this issue, their approach to synthesis of qualitative evidence began a new and major methodological line of inquiry (Thorne et al., 2004). While qualitative meta-synthesis methods have grown and developed in various directions, there are several common elements. The most basic is that qualitative meta-synthesis is both ‘hermeneutic’ and ‘dialectical’—or engaged in extracting and reporting constructs within studies, but also in comparing and contrasting between studies to understand variation and commonality (Jensen & Allen, 1996; Zimmer, 2006). Relatedly, qualitative meta-synthesis, like the qualitative research underlying it, is ‘interpretive’ rather than ‘aggregative’—that is, offering new interpretations and insights instead of combining studies, as in quantitative meta-analysis (Jensen & Allen, 1996; Thorne et al., 2004; Walsh & Downe, 2005). Sandelowski refers to the type of study that would vote-count or otherwise quantify qualitative research as ‘metastudy’ rather than true qualitative meta-synthesis (Thorne et al., 2004).

Another common element is the ‘reciprocal translation’, which, conceptually, is the understanding of one study’s findings in terms of other studies’ findings in an effort to develop syntheses that are consistent across included studies (Jensen & Allen, 1996; Walsh & Downe, 2005). A related idea, described in Noblit and Hare’s (1988) approach, is the ‘third-order construction’. This is the meta-ethnographic equivalent to the second-order construction that is a finding in primary qualitative research. These third-order constructions are the outcome of the qualitative meta-synthesis, in the way that second-order constructions are the outcome of the original qualitative report, and first-order

constructions are interpretations that the original study informants provide of their experiences (Malpass et al., 2009).

In developing syntheses of qualitative studies, Noblit and Hare (1988) described three main outcomes growing from an initial reciprocal translation. The first is 'reciprocal synthesis', which aims to understand larger constructs expressed by the studies in the inclusion set. A reciprocal synthesis arises when the second-order constructions expressed in studies broadly agree, and can be understood in terms of each other to form third-order constructions. In contrast, a 'refutational synthesis', which exposes and develops heterogeneity in the included studies, arises when inconsistencies between metaphors expressed in different studies contradict insurmountably and cannot be reciprocally translated. Finally, 'lines-of-argument synthesis' constructs a new, mid-range theory from the findings expressed in the included studies, and is used when studies report 'parts of the whole' phenomenon under investigation. Of these three approaches to meta-ethnography, Noblit reports that the first is most common, while refutational synthesis remains underutilised (Thorne et al., 2004).

Key philosophical and practical debates

Much of the resistance against collating qualitative research stems from paradigmatic differences, although there are both epistemological and practical reasons complicating the accumulation of qualitative findings (Britten et al., 2002). It is recognised that there are multiple and varying philosophies within the interpretivist paradigm, which impact qualitative studies to a larger degree than quantitative studies rooted nearer the positivist end of the paradigm spectrum (Dixon-Woods, Agarwal, Young, Jones, & Sutton, 2004). Thus, the critical question arises of *if* it is appropriate to synthesise across qualitative findings and if so, *how* (Atkins et al., 2008).

Those who would argue against the appropriateness of meta-synthesising qualitative research have pointed to the highly contextualised nature of qualitative findings and concern of watering down individual insights in favour of a generic summary (Sandelowski et al., 1997). Reducing qualitative studies into a summary would indeed be rejected by those who are firmly situated in the interpretivist paradigm on the basis that there is no one underlying truth (Thorne et al., 2004) and that searching for such a universal reality breaches core interpretivist tenets of qualitative research (Zimmer, 2006). Another question arising from this field asks whether qualitative meta-synthesis should synthesise across approaches, such as phenomenology or grounded theory (Jensen &

Allen, 1996)? One possible approach is to carry out simultaneous syntheses across studies grouped by framework and then 'triangulate' to validate findings (Finfgeld-Connett, 2010). Other methods, such as meta-study and meta-narrative, take these differences in methods as foci for analysis in their own right (Greenhalgh et al., 2005; Paterson, Thorne, Canam, & Jillings, 2001). Finally, Zimmer (2006) advances that with due attention to the epistemological differences underlying qualitative methodologies, synthesis across qualitative traditions is acceptable and appropriate.

A major issue on which Sandelowski and Barroso (2007) take a clear position is whether qualitative meta-syntheses should be based on purposive or comprehensive samples of studies. Though Thomas and Harden (2008) clearly assert the value of systematic search and retrieval in generating a qualitative meta-synthesis, Sandelowski and colleagues (Sandelowski, 2012; Sandelowski & Barroso, 2007; Sandelowski et al., 1997) state that a purposive sample is appropriate, as this is the same logic that guides primary qualitative research. However, because qualitative reports are themselves partly interpretations and partly summaries, it is an open question whether a purposive sample could truly lead to theoretical saturation, especially when nuances between themes and the relationships between themes expressed in different studies could continue to inform analysis in ways not otherwise detected through purposive sampling.

Moreover, application of primary qualitative data sampling principles such as 'saturation' does not quite translate to a meta-synthesis context, because the nature of published studies is to add something new to the literature base, suggesting that the data (i.e., each study) should diverge rather than converge (Dixon-Woods et al., 2006). The question remains, then, of how much data is sufficient when conducting a qualitative meta-synthesis (Jensen & Allen, 1996). Practically speaking, sampling primary studies is different from sampling empirical cases. There are a finite number of the former but not the latter, and there may be more pronounced heterogeneity between the former (e.g. between the foci of different studies) than the latter (e.g. between different study participants unified by a common set of characteristics). Even in some qualitative meta-syntheses where studies do largely 'converge' rather than 'diverge', it might not be possible to achieve saturation due to small numbers of included studies.

Furthermore, consideration of the quality of individual qualitative studies raises additional questions, as the practice of critically appraising qualitative research has reached no consensus on how it should be accomplished (Britten et al., 2002; Chapple &

Rogers, 1998). Should qualitative studies be excluded based on quality, and how is quality appraised (Walsh & Downe, 2005)? One argument posits that the inability to define a 'unified' qualitative research paradigm means that each study should be judged on its own merits, rather than on one set of guidelines (Rolfe, 2006). Conversely, Dixon-Woods, Shaw, Agarwal, and Smith (2004) posit that certain characteristics of qualitative research reporting and conduct are constant, regardless of methodology. Sandelowski and Barroso (2007) recommend appraisal, while advising against excluding studies based on perceived quality. Thus, including or excluding studies on the basis of quality is understandably contentious, although assessment of quality can become a focus for analysis. In a quantitative meta-analysis, researchers utilise sensitivity analyses to test for any influence of studies deemed to be of lesser quality on those meeting the criteria for high quality research. However, the process by which researchers might conduct sensitivity analyses on qualitative data is less certain (Dixon-Woods et al., 2006). Another approach is to rely on a study's peer-review-status as indication of quality control (see Boxes C and D). While the peer-review and revision process serves to check quality, including only peer-reviewed studies does not address findings from non-peer-reviewed reports, which might be of similar quality.

Along a more practical vein, empirical studies testing the use of thesaurus terms versus free-text terms versus broad-based terms when applying the same computerised search strategies commonly used in quantitative reviews to a qualitative context demonstrate such searches are not similarly productive (Shaw et al., 2004). It is the case that often qualitative studies are published outside of peer-reviewed journals registered with electronic databases, and that when they are in such journals they are not easily retrieved due to how they are abstracted (Green & Thorogood, 2014). All of this indicates that conducting a systematic search for qualitative studies is not as straightforward compared to similar searches for quantitative trial studies. However, it is important to remember that when searching electronic databases, search strategies can be tailored to favour sensitivity or precision, and electronic databases will vary in the sensitivity and precision of the articles they retrieve for any given search strategy (Taylor, Wylie, Dempster, & Donnelly, 2007). Furthermore, search strategies are continuously evolving and a recent systematic review on text mining for identifying relevant studies suggests that there are benefits to applying this technique to prioritise or reduce the number of studies considered for a review, which could allow researchers to favour a broader search strategy

without losing valuable time or requiring additional individuals to screen potential studies (O'Mara-Eves, Thomas, McNaught, Miwa, & Ananiadou, 2015).

Finally, is qualitative meta-synthesis intended to be a practice-relevant endeavour, particularly given Noblit's later reservations regarding its use in the clinical paradigms of evidence-based practice (Thorne et al., 2004)? Given the subjectivity and reflexivity inherent in all qualitative research enterprises, it is an open, inadequately addressed question in the application of meta-synthetic knowledge to clinical practice how the positivist and interpretive paradigms meet, if not reconcile, to allow knowledge transfer from one to the other (Paterson et al., 2001; Zimmer, 2006).

Methodological issues in practice

The diversity of opinion on qualitative meta-synthesis design is mirrored in its practice. Even in those papers embracing Noblit and Hare (1988), many approaches exist. Researchers have found the idea of reciprocal translation difficult to put into practice (Atkins et al., 2008). For example, Tondeur et al. (2012) conducted a reciprocal translation of qualitative studies describing strategies for integrating technology into the classroom (see Box A). They arranged studies chronologically, with the first paper identified as their 'index paper', or starting point. Then, they synthesised the first two papers by 'comparing the themes from paper one with paper two' (Tondeur et al., 2012, p. 136). This synthesis was used to study the third paper, with each subsequent paper being included in the synthesis one-by-one. The outcome was a set of third-order constructions that described a larger interpretive framework (Noblit & Hare, 1988). However, the authors describe the difficulties of simplifying and reducing the complexity found in the primary qualitative studies during translation (Tondeur et al., 2012).

[Box A near here.]

Jamal et al. (2013) describe a clearer translation process (see Box B). After rating each study for its conceptual richness and grouping the studies according to specific health topics, they identified high quality papers as 'index papers' under each topic grouping, and spiralled the reciprocal translation out from these points. For each of the two syntheses, the reciprocal translation was then developed by checking how papers' second-order constructions mapped onto the other papers in the set, and then taking those second-order constructions with the most explanatory power—that is, those second-order constructions that best mapped across included studies—to develop the reciprocal translation. Jamal et

al. (2013) then extend their analysis to create a lines-of-argument synthesis arising from these reciprocal translations.

[Box B near here.]

In sum, even those studies taking a meta-ethnographic approach Noblit and Hare (1988) demonstrate various approaches to reciprocal translation and meta-synthesis. These critiques of both the concept and processes of systematically reviewing and meta-synthesising qualitative research must be weighed against those directed towards the current field of qualitative research existing in isolation without a cumulative knowledge base. Green and Thorogood (2014) explain two strengths of systematically reviewing qualitative research to include: (1) identifying common themes across contexts providing the opportunity to discuss the generalisability of some findings (typically not a goal of primary qualitative research) and (2) assessing how individual qualitative methods impact the amount and type of data generated in primary qualitative studies.

Other researchers in favour of advancing meta-synthesis methods have argued that to ‘claim that generalisation is not possible is to deny the transferability of any shared meanings or generative mechanisms’ (Britten et al., 2002, p. p. 214), and that the findings from qualitative meta-syntheses offer new insights, which are greater than the sum of their parts (Campbell et al., 2003). Indeed, proponents of meta-synthesis maintain that systematically reviewing and meta-synthesising qualitative findings advances both the qualitative and quantitative research fields in a number of relevant ways, including: generating mid-range theory (Estabrooks et al., 1994), developing concepts (Campbell et al., 2003), and adding range and complexity to effectiveness studies (Atkins et al., 2008). Furthermore, Brown and Lan (2015) utilise qualitative meta-synthesis to examine how national educational policies may have influenced educators’ perspectives of ‘school readiness’, and their conclusions have policy-relevant implications (see Box C).

[Box C near here.]

Methods for qualitative meta-synthesis

This discussion now turns its attention to demonstrating the large diversity in meta-synthetic approaches manifest in the literature. Rather than cataloguing different approaches which can be found elsewhere (Barnett-Page & Thomas, 2009; Dixon-Woods, Agarwal, et al., 2004), this discussion demonstrates the need for an assessment of current metasynthetic practice.

Several different approaches to qualitative meta-synthesis have emerged, with most connected to the procedures Noblit and Hare (1988) outlined originally. Atkins and colleagues (2008) interpreted the steps Noblit and Hare laid down as beginning with a research question; locating, filtering, and appraising reports of qualitative studies; ‘reading the studies’; understanding interrelations between studies; developing ‘reciprocal translations’ between studies; and then ‘synthesizing’ the studies and ‘expressing’ this synthesis.

Of note is that Noblit and Hare (1988) demonstrate the development of reciprocal translations and the synthesis of these translations, but they offer little guidance as to how to accomplish these steps. At an early point in the methodological development of meta-ethnographic methods, it appeared that the different approaches, like Noblit and Hare’s method, tended to remain vague in the details of their execution (Sandelowski et al., 1997). However, Sandelowski and colleagues (1997) laid down what they believed to be several principles, including purposive sampling of studies, meta-synthesis only of studies rooted in the same qualitative epistemology, and a target sample of about 10 studies (see Box D). [Box D near here.]

An important variant of the Noblit and Hare (1988) approach is ‘critical interpretive synthesis’ (Dixon-Woods et al., 2006). Dixon-Woods and colleagues begin by rejecting the reciprocal translation as adequate for capturing the complexity inherent in collections of qualitative research. Instead, they adopt the lines-of-argument synthesis, arguing that qualitative meta-synthesis should go beyond reciprocal translation to be both critical—engaged in how researchers construct the interpretations of the phenomenon—and interpretive, or engaged in how the interpretations offered by researchers compare and contribute to higher order understandings. They describe the output of their meta-synthesis as the ‘synthetic argument’, a term clearly connected with the conceptual heritage of the ‘lines-of-argument synthesis’ (Barnett-Page & Thomas, 2009; Dixon-Woods et al., 2006; Noblit & Hare, 1988).

Another major approach to qualitative meta-synthesis called the ‘meta-study’ describes a process simultaneously meta-synthesising and reciprocally translating methods, results, and discussion to understand how these three sections interact and form mid-range theory relating to the object of study (Paterson et al., 2001). It is noted that this version of ‘meta-study’ is different from the ‘meta-study’ which Sandelowski describes elsewhere as primarily a vote-counting summary (Thorne et al., 2004). The ‘meta-methods’ analysis

elaborates the different methods, data, and epistemologies used to examine the phenomenon of interest (Barnett-Page & Thomas, 2009). Unlike Dixon-Woods and colleagues' (2006) critical interpretive synthesis, the meta-study's 'meta-methods' and 'meta-discussion' form distinct parts of the final analysis, where the parts are then integrated to view how each informs the other. Paterson and colleagues' meta-synthetic method shares much in common with 'meta-narrative', an emergent approach to research synthesis that concerns itself as well with how these different epistemologies shape results and understandings of a phenomenon. In particular, meta-narrative is used in mixed-methods reviews, in which quantitative and qualitative evidence are examined together (Greenhalgh et al., 2005).

Finally, one of the most recent approaches to qualitative meta-synthesis is Thomas and Harden's (2008) 'thematic analysis' approach, which marshals qualitative evidence for practice-based relevance and for intervention development and evaluation (see Box E). This method involves line-by-line coding of the findings of a study to develop specific hypotheses to inform systematic reviews and intervention development and without the articulated underpinnings of grounded theory.

[Box E near here.]

A grounded theory-inspired approach (see Box F) develops the meta-synthesis by constant comparison through line-by-line coding of study findings and recording analytic memos (Finfgeld-Connett, 2010). Finfgeld-Connett (2010) asserts that a well-developed meta-synthesis using these approaches to triangulate across studies with different qualitative methodologies, which she terms *meta-interpretation*, improves generalisability and transferability of the findings of the meta-synthesis, though the author does not further explain these benefits. Another grounded theory-inspired method of qualitative meta-synthesis emphasises the role of metaphors—the building block of qualitative analysis—in developing 'grounded formal theory' from the findings of qualitative studies (Kearney, 2001).

[Box F near here.]

Next steps for qualitative meta-synthesis

As meta-ethnography and qualitative meta-synthesis are relatively new, the development of these analytic approaches and extension of methods to new challenges are relatively diffuse. Much of the methodological work to advance this design has taken place in the health sciences. However, this presents an opportunity for educationalists to 'reclaim' and

develop meta-synthetic methods to meet their specific needs. We identify three key ways in which qualitative meta-synthesis is advancing as a design: in terms of the aspects of systematic review methodology needed to support successful meta-syntheses, in terms of the methods used to analyse qualitative studies and to use the results of that analysis, and in terms of the discursive, methodological and paradigmatic dimensions of qualitative meta-synthesis.

Laying the groundwork for successful meta-syntheses

One key way in which methods for qualitative meta-synthesis are developing is in respect of the systematic review methodology needed to set up meta-syntheses, search for relevant studies and report the outcomes of the meta-synthesis. One important recent contribution in this regard is the SPIDER tool to define key parameters for a qualitative meta-synthesis (Cooke, Smith, & Booth, 2012). Analogous to the population, intervention, comparison and outcome heuristic used to guide systematic reviews of intervention effectiveness, SPIDER stands for sample, phenomenon of interest, design, evaluation, and research type. Cooke and colleagues (2012) suggest that using the SPIDER tool is a more effective way to define the key 'parameters' for a systematic review of qualitative research.

Once the question has been set, the issue of how to search for studies presents itself. Systematic reviewers in the field of education have long acknowledged the challenges inherent in searching for qualitative research, due to the panoply of methods, approaches, and key terms, as well as the highly variable indexing of studies (Thomas & Harden, 2008). The development of specific guidelines to search for different types of studies to inform systematic reviews is most advanced for randomised trials—for example, the Cochrane highly sensitive search strategies for randomised trials (Higgins & Green, 2008)—but work on best practices in locating qualitative studies continues apace. Reflecting and analysing their experience in searching for qualitative studies to inform four different systematic reviews, Stansfield, Brunton, and Rees (2014) suggest that it is important to 'search wide and dig deep' by using a wide variety of databases and sources, including grey and fugitive literature, and by using sensitive search strategies. In this regard, it may be of utility for educationalists undertaking systematic reviews of qualitative research to consider searching health-oriented databases such as MEDLINE or PubMed and Embase, which are especially broad in their reach and well-indexed. However, qualitative research itself may not be well detected by subject headings corresponding to method (Bradley, 2012). Some meta-synthesists have rejected the idea that a one-stage

search is even appropriate for systematic reviews of qualitative research and that search strategies, while auditable, should be considered emergent and documented as such in reports. That is, systematic reviews of qualitative research should be iterative and driven by a ‘berry-picking’ approach (Finfgeld-Connett & Johnson, 2013), which mirrors snowballing sampling techniques in primary qualitative research. It should be noted that just as in primary qualitative research, a preliminary synthesis might lead researchers to change inclusion criteria, which would be unusual for quantitative syntheses.

Finally, complete and transparent reporting of qualitative meta-syntheses is important in the interest of both systematic review methodology and strong qualitative research practice (see Box B). The myriad guidelines for reporting other types of studies—such as PRISMA for systematic reviews of interventions (Liberati et al., 2009) and COREQ for qualitative research (Tong, Sainsbury, & Craig, 2007)—taken together with the unsuitability of other frameworks for reporting qualitative meta-syntheses, suggest that systematic reviews of qualitative research require a bespoke reporting guideline. That guideline has come in the form of enhancing transparency in reporting the synthesis of qualitative research, or ENTREQ (Tong, Flemming, McInnes, Oliver, & Craig, 2012). As qualitative meta-synthesis evolves, it is likely that reporting guidelines will continue to evolve as well.

Horses for courses and qualitative meta-synthesis

Previous reviews have identified the proliferation of methods for qualitative meta-synthesis (Barnett-Page & Thomas, 2009; Dixon-Woods, Agarwal, et al., 2004; Hannes & Macaitis, 2012; Melendez-Torres, Grant, & Bonell, 2015). However, methods continue to grow and develop as qualitative meta-synthesists innovate to meet their specific needs, and as meta-synthesists seek to overcome the qualitative-quantitative divide across different systematic reviews.

New analytic methods. Much of the work on development of new analytic methods has been in response to specific disciplinary or contextual issues that arise in doing a systematic review of qualitative research. To our knowledge, there has not been a specific method recently developed to meet the needs of educationalists, though of course the germinal work in meta-ethnography was conducted by a team of educational anthropologists (Noblit & Hare, 1988). However, recent work by social workers on ‘qualitative interpretive metasynthesis’ (Aguirre & Bolton, 2014) has been led by a need to translate the principles of qualitative meta-synthesis into the values of social work, for

example by focusing on ‘synergistic understandings’ that reflect social workers’ engagement with clients’ ecological systems as opposed to isolated themes arising from a meta-synthesis of qualitative studies. The importance of the role of ecological systems is by no means limited to the field of social work and is very relevant to educational researchers as they examine the multiple and intersecting environmental layers inherent in schools.

As an example of the development of methods to meet specific contextual issues, ‘best fit’ framework synthesis (Carroll, Booth, Leaviss, & Rick, 2013) builds on previous work in framework synthesis (Brunton, Oliver, Oliver, & Lorenc, 2006) to work through qualitative studies quickly and extend a priori conceptual frameworks in contexts where findings of a qualitative meta-synthesis are required to be responsive and quickly delivered. Carroll and colleagues (2013) simultaneously assemble relevant qualitative studies and theoretical frameworks, and they use a synthesis framework to code information from the primary studies. Data that cannot fit into the framework are then examined to nuance or extend the framework.

Putting systematic reviews together. An emerging approach in this field includes the ‘mixed methods systematic review’, which incorporates elements from both quantitative systematic reviews and meta-analyses as well as from qualitative meta-syntheses. Again, key instructive examples are found among health research—notably Harden et al. (2004) and Oliver et al. (2005)—rather than education, although the EPPI-Centre’s review library is notable for cataloguing research in this area. Overall, there remains a need to more comprehensively bridge the gap between qualitative and quantitative educational research in syntheses.

Such a cross-synthesis between quantitative and qualitative review findings may be conducted a number of ways, including according to a mixed methods convergent parallel design. The convergent parallel design involves a phase in which qualitative and quantitative data are concurrently yet independently collected and analysed, followed by a phase which sees these two strands joined together and the data merged for subsequent interpretation and discussion (Creswell & Plano Clark, 2011). Based in pragmatism, this design places the mixed methods nature of the research question at the centre, casting aside any controversies arising from paradigm differences between constructivism and positivism in favour of applying a practical approach to research (Creswell & Plano Clark, 2011). This design may be most appropriate to answer mixed methods research questions

because it allows for the synthesis of corresponding data across qualitative and quantitative results so as to more fully illuminate the phenomenon of interest, valuing each strand equally without placing one above the other (Creswell & Plano Clark, 2011). It is important to point out that while the linear nature of written reports may suggest that the strand reported first is conducted first and the strand reported second is conducted after the first, this is not the case in a convergent parallel design, where both strands are conducted in tandem.

Another design for cross-synthesising between qualitative and quantitative primary studies includes conducting an initial qualitative synthesis to develop theory and suggest hypotheses, which could then be tested in a quantitative synthesis. Similarly, the reverse order could also be illuminating, particularly if qualitative meta-syntheses were conducted to delve deeper into exploring findings from quantitative syntheses. It is important to bear in mind that whichever design is applied, there are distinctive strengths of both qualitative and quantitative research, and whichever design is selected for the mixed methods synthesis should reflect the nature of the research questions.

Discursive and methodological issues

A key area in which qualitative meta-synthesis is developing is in outlining the discursive issues it presents for social scientists, including educationalists. For example, Riese, Carlsen, and Glenton (2014) have recently argued that even in anthropologically oriented understandings of education, qualitative meta-synthesis can be helpful both to build up theoretical understandings across contexts and to problematise traditional understandings of a research field. Their argument, which essentially is oriented towards the development of an applied anthropology that is analytically generalisable, does not demand a specific analytic approach, aside from a general orientation towards interpretivism. Riese and colleagues (2014) also point out that as a method, systematic reviews of qualitative research can be used to set the agenda for future social research.

Within the operation of qualitative meta-synthesis, methodologists continue to seek clarity on ‘how’ studies are synthesised. This bears obvious parallels to continue innovation in analysis of primary qualitative data. Melendez-Torres, Grant and Bonell (2015) examined a set of 61 qualitative meta-syntheses to establish a taxonomy of how these meta-syntheses undertook to achieve reciprocal translation. They divided the approaches used into four categories: one in which reciprocal translation was based on a visual representation, such as a map of concepts or a grid of second-order constructs; one

in which reciprocal translation began with a ‘key paper’ and then examined subsequent papers to develop the synthesis (see Boxes A and B); one in which study themes were used as the basis of the reciprocal translation either by combining them into similar categories or by comparing them across studies; and one in which meta-synthesists used an approach similar to primary qualitative research in coding study results line by line. This taxonomy, though developed in public health, offers a parsimonious classification to understand the diversity of approaches in qualitative meta-synthesis. In addition to this taxonomy of operations of reciprocal translation, one research group has described that meta-ethnography involves two processes, the first of which is ‘reading’, or engagement with the text of the studies ‘to appraise; familiarise; identify; extract; record; organise; compare; relate; map; stimulate; and verify’ (Lee, Hart, Watson, & Rapley, 2014, p. 8), and the second of which is ‘conceptual innovation’, or the development of key metaphors that go beyond the primary studies, rather than merely restating them. Lee and colleagues (2014) point out that conceptual innovation is not inevitable in meta-ethnographies, due in part to the path-dependency of research narratives.

Conclusion

This paper has reviewed paradigmatic and methodological aspects of qualitative research before exploring the historical movement, core theoretical and practical issues of qualitative meta-synthesis. The spectrum of methods reflects the different approaches to qualitative research visible at the primary study level. As such, in order for meaningful progress in this area to contribute to educational practices, researchers must be both transparent and comprehensive in their reporting of chosen meta-synthesis aims and methods. The opportunity to apply qualitative meta-synthesis alongside quantitative systematic reviews and meta-analyses that are already more widely accepted in the education field opens up future discovery of novel insights for theory development as well as everyday practice and research agendas. Given the multiple ecological levels at which educational policies and practices are initiated in addition to the recognition of children’s individual learning needs, much can be gained by a broader uptake of qualitative meta-synthesis and in turn mixed methods cross-syntheses within the field of educational research.

Sponsorship

None to declare.

Box A. Preparing pre-service teachers to integrate technology in education: a synthesis of qualitative evidence (Tondeur et al., 2012)

In their 2012 paper on integrating technology in education, Tondeur and colleagues aimed to identify strategies for instructing pre-service teachers on how to incorporate technology in their lessons. To do this, the authors conducted a key word search in the electronic database Web of Science, which while not fully elaborated in the paper, somehow resulted in 144 studies to be assessed initially by reading their abstracts for relevance to this meta-synthesis.

The 23 studies that passed this examination were then critically appraised using the Critical Appraisal Skills Programme (CASP) checklist for qualitative research (1988). Although potential included studies were systematically and rigorously appraised, the authors did not exclude studies based on poor ratings or lack of comprehensively reported methodologies. This quality appraisal was explicitly included as the third phase of Tondeur and colleagues' meta-ethnography, which otherwise closely aligned with Noblit and Hare's original (1988) description. The authors chose to chronologically translate each study into the next, such that the first paper to be published served as the 'index paper' whose themes were compared with those of the second paper to be published as so forth. Throughout this process, the concepts were reduced and abstracted by 'merging and collapsing' (p. 136) themes of similar groupings. The authors flagged such reduction as a potential limitation of their study in that they acknowledge the 'reduced...importance of the contextuality of the results...the results of the review cannot simply be generalized to other teacher training institutions in different parts of the world' (p. 142).

The findings of this meta-synthesis are reported at both a micro and institutional level. While the authors do not go so far as to explicitly provide guidance to practitioners based on their findings, they do assert, '...it seems to be important that pre-service teachers have the possibility to see and experience the pedagogical integration of technology in the classroom during their training experiences' (Tondeur et al., 2012, p. p. 142). Additionally, they argue that unless theory is delivered alongside practical instruction, 'the knowledge and the skills pre-service teachers gained from a separate stand-alone course process are likely to remain isolated and unused' (p. 142). Thus, although couched in their report as not generalisable to other contexts, the authors do express a desire that the themes from this meta-synthesis inform policy makers and

practitioners in this area.

Box B. The school environment and student health: a systematic review and meta-ethnography of qualitative research (Jamal et al., 2013)

In their 2013 meta-ethnography, Jamal and colleagues examined the processes through which young people's health was influenced by their social and physical school environment. In this thoroughly reported study, readers are directed to the previously published study protocol for additional specifics regarding search strategy and terms etc. (Bonell et al., 2011). The practice of publishing a protocol provides greater transparency in the research field regarding intended aims and proposed analyses, which can guard against reporting or publication bias given undesirable or underwhelming findings. While typically more discussed among a quantitative tradition (e.g., randomised controlled trials and systematic reviews of trials), this practice in qualitative meta-synthesis allows for much greater coverage and depth of methodological descriptions in the protocol and thus more room to report interpretations and findings in the main paper.

Indeed, this meta-synthesis sought out studies with thick descriptions by rating potential included studies as 'high', 'medium', or 'low' in conceptual richness. The authors additionally applied criteria from EPPI-Centre guidelines to appraise the quality of included studies. While they did not exclude studies with lower ratings, Jamal and colleagues used these ratings to inform their interpretation, using studies deemed to be of higher quality as the starting point or 'index papers' in their meta-synthesis.

Unlike the other meta-syntheses described in this paper, Jamal and colleagues created an evidence map following their initial search for relevant studies. They took this map to key stakeholders (e.g., young people, policy makers) for consultation and setting priorities of their research agenda. In this way, we see how the exchange of knowledge was bidirectional between the research team and stakeholders. Additionally, this process provides an example of how an *a priori* protocol need not stymie the exploratory and inductive process of qualitative research.

When translating the themes and concepts from one study into the others, Jamal and colleagues first grouped the studies based on the type of health outcome reported in the primary study. After identifying key themes within each outcome type, relationships between these themes/outcomes were established. Finally, as their focus was on the processes of school environment on students' health, the authors established their line of

argument interpretation. The authors note how young people sought to create and maintain a ‘tough’ reputation, as this would allow them entry into friend groups with other ‘tough’ young people, thus providing safety. Understandings of masculinity in the school environment were reflected in this ‘toughness’. Furthermore, ‘unowned’ (p. 6) spaces such as staircases and toilets provided opportunities for risky health behaviours compared to classrooms, which were associated with oversight provided by the class teacher. In addition, young people reported that their teachers were ‘disconnected’ (p. 6) from their day-to-day realities, particularly from those young people whose home lives were particularly unstable or those who identify as Black. Finally, because of their disconnection from school, young people sought to ‘escape’ (p. 7) the school area when possible, such as during lunch periods, which in turn saw them purchasing unhealthy foods and beverages from local shops or smoking. Thus, Jamal and colleagues reached the conclusion that students ‘not only react to schools’ institutional systems for ordering instructional and regulatory practices, but they also promote their own parallel, competing version of these...‘orders’...where students’ family and/or community culture is immersed in urban ‘street culture’, with relatively little hope of conventional social advancement, this will permeate the local student-network and thus shape both students’ action and, in turn, the institutions’ regulatory response’ (2012, p. 8). It is clear, then, that this meta-synthesis extended beyond describing the themes found in the primary studies to posit a theory by which students and educational institutions act and react to each other in ways that influence health decisions.

Box C. A qualitative metasynthesis comparing U.S. teachers’ conceptions of school readiness prior to and after the implementation of NCLB (Brown & Lan, 2015)

This 2015 paper by Brown and Lan focuses on teachers’ notions of school readiness and whether/how their understandings changed with the implementation of the No Child Left Behind (NCLB) Act in the United States. They searched electronic databases and conducted key-word searches in relevant journals, although a comprehensive list of databases and journals is not provided as in some of the other studies highlighted here as examples. It is similarly unclear how many references were generated by the initial searches and what the procedure was for narrowing down this initial list to result in their first 38 potential studies. In applying a quality control to these potential included studies, Brown and Lan chose the criterion of publication in a peer-reviewed journal as indication

that the reported study was of sufficient quality for inclusion in the meta-synthesis. Their reasoning was that in the process of submission, peer review, revision, and ultimate publication, any study should have been rigorously assessed by academic peers and thus its trustworthiness would be higher than reports in non-peer-reviewed publications. By excluding studies not published in peer-reviewed journals, the authors further purged their potential included studies to 17. The authors do acknowledge the potential limitation of this form of quality control, writing that ‘it might have such unintended consequences as filtering out research findings that contradict the findings of this metasynthesis’ (Brown & Lan, 2015, p. 4). Given that one of the three types of meta-ethnography described by Noblit and Hare (1988) is that of refutational hypothesis, and that such refutational meta-syntheses are more rare in the literature than reciprocal translations or lines of argument syntheses, this limitation is well noted. Their final set of included studies numbered 12 (six pre-NCLB and six post-NCLB), as they sought to purposively restrict their included studies to between 6 and 10 papers, as per guidelines by (Major & Savin-Baden, 2010).

Based in meta-ethnographic roots, Brown and Lan name their process as qualitative meta-synthesis and follow the procedures outlined in Major and Savin-Baden’s (2010) book on qualitative research synthesis. Whereas Noblit and Hare’s (1988) meta-ethnography contains seven phases, Brown and Lan modelled their qualitative meta-synthesis off of the four-step process described by Major and Savin-Baden (2010). In this process, which can be roughly mapped on to meta-ethnography, the first step ‘beginning the process’ entails phases one through four of meta-ethnography, followed by step two ‘analysis’ (phase five), step three ‘synthesis’ (phase six), and step four ‘interpretation’ (phase seven). During the analysis and synthesis stages, Brown and Lan identified initial codes from the studies that led to the development of primary and ultimately secondary themes.

The findings of this meta-synthesis conclude that there was indeed a shift in how teachers’ constructed meaning of ‘school readiness’ with the implementation of the NCLB Act. The authors posit that increasing focus on academic outcomes will intensify this shift between nativism (school readiness as a within-child concept, i.e., maturity) towards empiricism (school readiness as the result of preschool teachers’ work to prepare children), even as it perpetuates a fairly limited, White middle-class perception of what it means for children to be ready for school. Thus, it appears that with this shift in attitudes

about school readiness, teachers increasingly view it as their responsibility to train preschool-aged children on expectations aligned with White, middle-class values for academic/social skills. Brown and Lan argue that from these findings, more work needs to occur when training teachers to ‘understand the messiness of the construct of school readiness’ (2015, p. 10). Thus, the findings from their meta-synthesis are intended if not to direct *how* practice needs to change then to at least serve as *a call* for change in practice and policy.

Box D. Research on mentor education for mentors of newly qualified teachers: a qualitative meta-synthesis (Aspfors & Fransson, 2015)

In their 2015 paper on educating mentors for newly qualified teachers, Aspfors and Fransson sought to identify common themes across studies on mentor education as well as insights into how to advance mentor education as a field. Aspfors and Fransson widened their search for relevant studies beyond just that of electronic databases. They also searched peer-reviewed journals and perused reference lists of included studies. While systematic reviews of quantitative studies aim for comprehensiveness when sampling, it is yet to be universally decided whether reviews of qualitative research should also seek an all-inclusive sample or if primary qualitative research sampling techniques such as purposive sampling are appropriate. In this example, Aspfors and Fransson follow recommendations along the purposive sampling argument and actively seek 10 to 12 included studies for their meta-synthesis (Bondas & Hall, 2007 as cited in Aspfors & Fransson, 2015). In addition to the inclusion criteria related to their topic, the authors used publication in a peer-review journal as a proxy for assessment of high quality research, which mirrors the example in Box C. Thus, studies not published as peer-review articles were excluded on the grounds of lack of quality.

Aspfors and Fransson based their analysis on an adapted meta-ethnography, which is also based in Sandelowski and Barroso’s (2007) qualitative research synthesis. For this process, when they were initially relating the themes from different included studies, the authors applied qualitative content analysis, in which they proceeded from open coding to categorisation across codes to a yet higher level of abstraction. This resulted in a meta-synthesis that produced three broad dimensions within which the authors situated aspects of mentor education: (1) the educational context (e.g., national educational policy); (2) the interplay of theory and practice, which ‘are inseparable units,

where one cannot be understood without the other' (Aspfors & Fransson, 2015, p. 83); and (3) the relationship between mentors as a way to build an open culture for trust. More specifically, the researchers were able to identify common themes, including the importance of the school context in facilitating the mentor education, the extent to which mentors were able to concurrently practice the theoretical concepts or relate the concepts to previous practice, the benefits of self-reflection, and the role of positive relationships between mentor and mentee. These findings, more than just describing the concepts of the primary qualitative studies, identify overarching dimensions that the authors posit are necessary for consideration when establishing and implementing mentor education programmes.

Box E. Barriers to, and facilitators of, parenting programmes for childhood behaviour problems: a qualitative synthesis of studies of parents' and professionals' perceptions (Koerting et al., 2013)

Koerting and colleagues focused on improving children's home learning environment through examining barriers and facilitators to parent training programmes in their 2013 meta-synthesis. The authors studied aspects related to both initial access to these parent programmes as well as continued participation, and they sought views from either parents attending the programmes or professionals facilitating the trainings. In addition to their search of electronic databases, which was elaborated in an appendix, Koerting and colleagues highlighted the necessity of performing additional web-based searches and perusing reference lists of included studies to avoid overlooking relevant reports that might not have surfaced in the original database search. These search methods are common among more traditional systematic reviews of quantitative studies, but are perhaps even more relevant for reviews of qualitative research given the previously discussed practical limitations of how qualitative research studies are catalogued online.

As with some of the other examples provided here, Koerting and colleagues applied EPPI-Centre criteria when appraising the quality of potential included studies. Unlike some other studies, the authors did exclude a paper based on a determination of poor quality; however, the specific reasons that paper was deemed to be poor quality are not elaborated. While even the process of appraising the quality of qualitative studies can be controversial, it has been shown thus far that educational meta-syntheses often do

conduct some type of quality appraisal. However, more variable is the decision of whether to exclude studies, as done in this meta-synthesis, for poor quality.

Koerting and colleagues applied thematic synthesis in this meta-synthesis, resulting in findings that, while still interpretive, are much more integrative than those reported in other examples following a meta-ethnographic approach. After coding each line of findings in the primary studies, the authors grouped these codes first into descriptive categories and subsequently into analytical themes. After working and re-working the multiple layers of codes, the authors went back to the original studies and applied their new framework when re-coding the findings sections. The result is a four-part description of the barriers/facilitators to initial access/continued participation, which are discussed in a preliminary way as moderated by type of respondent (e.g., parents or professionals). The authors were able to identify differences between parents' and professionals' views, with parents indicating concerns about barriers to service access and continued engagement, often with a focus on difficulties balancing their work responsibilities and caring for other children in the home. Professionals, in contrast, provided more views on both barriers and facilitators, but their concerns revolved more around challenges in provision of services. Since the design of this meta-synthesis was predominantly descriptive and utilised thematic synthesis methods, the interpretation does not extend to positing a larger framework of how these processes interact or inform practice.

Box F. Peer relations in peer learning (Riese, Samara, & Lillejord, 2012)

Riese and colleagues, in their 2012 paper on students' peer learning, sought to explore both the social and relational processes of peer learning in the educational environment through meta-synthesising primary qualitative research with thick descriptions. They gathered qualitative research examples through a systematic electronic database search, detailed in their report. They applied three criteria against which they judged potential included studies: (1) studies should have been carried out in a school setting with a teacher facilitating peer learning with such significant student collaboration that created and maintained relationships between peers; (2) studies should have explored the processes of peer learning and relations that are formed between the students; and (3) studies should have produced thick descriptions that could be compared with those of other studies (i.e., richly analysed and sufficiently reported). In addition to these

inclusion criteria, Riese and colleagues appraised the quality of included studies following guidelines advanced by Elliott, Fischer, and Rennie (1999), which they include in an appendix (Riese et al., 2012). While it is unclear whether studies were excluded based on the result of this quality assessment, it is clear that studies deemed to be insufficiently developed with regard to the thickness or richness of their descriptions were indeed excluded from this meta-synthesis on peer learning.

In analysing their included studies, Riese and colleagues followed a meta-ethnographic approach in which they apply grounded theory methods to examine the processes involved in peer learning. To this end, their analysis involved three stages of increasing levels of abstraction, as the first sought to identify all themes in the primary studies, then to compare those themes across studies, and finally to find novel insights. In response to critiques of meta-synthesis designs and the many methods researchers might use, the authors clarify that, 'Rather than considering the included studies as cases to be compared, we regard them as *texts* in a scientific discourse on peer learning... This implies an understanding of theory as identification of social rules that underlie and govern social patterns, rather than as law-like expressions of the nature of social behaviour' (Riese et al., 2012, p. 608). Thus, they emphasise that they interpreted the findings of the primary studies rather than the original data from those studies. It is in this way that they come to the new understandings termed 'third order interpretations' in the final step of their analysis.

Finally, Riese and colleagues directly link their new, meta-synthesised interpretation of peer learning to applicability for practitioners (i.e., teachers) when they provide guidance for *how* educators should go about integrating peer learning into the classroom. They write, '...it is of crucial importance to pay attention to the social relations that exist between members of a group prior to the design of the peer learning programme and in the development of the programme. In peer group interaction, peer relations must be perceived as integral to the design and instructional practices of the peer learning programme as they mediate action in both productive and less productive manners' (Riese et al., 2012, p. 618). Thus, this meta-synthesis extends beyond a thematic description of different types of peer learning processes and not only posits a theory for the mechanisms by which peer learning is or is not effective, but also advises practitioners on how to set up peer learning in a way to encourage beneficial results.

References

- Aguirre, R. T., & Bolton, K. W. (2014). Qualitative interpretive meta-synthesis in social work research: Uncharted territory. *Journal of Social Work, 14*(3), 279-294.
- Annells, M. (1997). Grounded theory method, part I: Within the five moments of qualitative research. *Nursing inquiry, 4*(2), 120-129.
- Aspfors, J., & Fransson, G. (2015). Research on mentor education for mentors of newly qualified teachers: A qualitative meta-synthesis. *Teaching and Teacher Education, 48*, 75-86.
- Atkins, S., Lewin, S., Smith, H., Engel, M., Fretheim, A., & Volmink, J. (2008). Conducting a meta-ethnography of qualitative literature: lessons learnt. *BMC medical research methodology, 8*(1), 21.
- Barnett-Page, E., & Thomas, J. (2009). Methods for the synthesis of qualitative research: a critical review. *BMC Medical Research Methodology, 9*(1), 59.
- Barroso, J., Gollop, C. J., Sandelowski, M., Meynell, J., Pearce, P. F., & Collins, L. J. (2003). The challenges of searching for and retrieving qualitative studies. *Western journal of nursing research, 25*(2), 153-178.
- Bonell, C., Harden, A., Wells, H., Jamal, F., Fletcher, A., Petticrew, M., . . . Murphy, S. (2011). Protocol for a systematic review of the effects of schools and school-environment interventions on health: evidence mapping and syntheses. *BMC public health, 11*(1), 453.
- Bowling, A. (2009). *Research methods in health: Investigating health and health services* (3rd ed.). Maidenhead: Open University Press.
- Bradley, S. M. (2012). Examination of the Clinical Queries and Systematic Review “hedges” in EMBASE and MEDLINE. *Journal of the Canadian Health Libraries Association, 33*(1), 34.
- Britten, N., Campbell, R., Pope, C., Donovan, J., Morgan, M., & Pill, R. (2002). Using meta ethnography to synthesise qualitative research: a worked example. *Journal of Health Services Research & Policy, 7*(4), 209-215.
- Brown, C. P., & Lan, Y.-C. (2015). A qualitative metasynthesis comparing US teachers' conceptions of school readiness prior to and after the implementation of NCLB. *Teaching and Teacher Education, 45*, 1-13.

- Brunton, G., Oliver, S., Oliver, K., & Lorenc, T. (2006). A Synthesis of Research Addressing Children's, Young People's and Parents' Views of Walking and Cycling for Transport.
- Campbell, R., Pound, P., Pope, C., Britten, N., Pill, R., Morgan, M., & Donovan, J. (2003). Evaluating meta-ethnography: a synthesis of qualitative research on lay experiences of diabetes and diabetes care. *Social science & medicine*, 56(4), 671-684.
- Carroll, C., Booth, A., Leaviss, J., & Rick, J. (2013). "Best fit" framework synthesis: refining the method. *BMC Medical Research Methodology*, 13(1), 37.
- Chapple, A., & Rogers, A. (1998). Explicit guidelines for qualitative research: a step in the right direction, a defence of the 'soft' option, or a form of sociological imperialism? *Family practice*, 15(6), 556-561.
- Charmaz, K. (2006). Constructing grounded theory: A practical guide through qualitative research. *Sage Publications Ltd, London*.
- Cooke, A., Smith, D., & Booth, A. (2012). Beyond PICO The SPIDER Tool for Qualitative Evidence Synthesis. *Qualitative Health Research*, 22(10), 1435-1443.
- Corbin, J., & Strauss, A. (Eds.). (2008). *Basics of qualitative research: Techniques and procedures for developing grounded theory*. London: Sage.
- Creswell, J. W., & Plano Clark, V. L. (2011). *Designing and conducting mixed methods research* (2nd ed.). Thousand Oaks, CA: SAGE.
- Critical Appraisal Skills Programme (CASP). (1988). *Ten questions to help you make sense of qualitative research*. Oxford: CASP.
- Dixon-Woods, M., Agarwal, S., Young, B., Jones, D., & Sutton, A. (2004). *Integrative approaches to qualitative and quantitative evidence*: Health Development Agency London.
- Dixon-Woods, M., Bonas, S., Booth, A., Jones, D. R., Miller, T., Sutton, A. J., . . . Young, B. (2006). How can systematic reviews incorporate qualitative research? A critical perspective. *Qualitative Research*, 6(1), 27-44.
- Dixon-Woods, M., Shaw, R. L., Agarwal, S., & Smith, J. A. (2004). The problem of appraising qualitative research. *Quality and Safety in Health Care*, 13(3), 223-225.
- Dixon- Woods, M., Fitzpatrick, R., & Roberts, K. (2001). Including qualitative research in systematic reviews: opportunities and problems. *Journal of Evaluation in Clinical Practice*, 7(2), 125-133.

- Elliott, R., Fischer, C. T., & Rennie, D. L. (1999). Evolving guidelines for publication of qualitative research studies in psychology and related fields. *British Journal of Clinical Psychology*, 38(3), 215-229.
- Estabrooks, C. A., Field, P. A., & Morse, J. M. (1994). Aggregating qualitative findings: an approach to theory development. *Qualitative health research*, 4(4), 503-511.
- Evans, D. (2002). Database searches for qualitative research. *Journal of the Medical Library Association*, 90(3), 290-293.
- Fine, G. (1993). The sad demise, mysterious disappearance, and glorious triumph of symbolic interactionism. *Annual Review of Sociology*, 19, 61-87.
- Finfgeld- Connett, D. (2010). Generalizability and transferability of meta- synthesis research findings. *Journal of Advanced Nursing*, 66(2), 246-254.
- Finfgeld- Connett, D., & Johnson, E. D. (2013). Literature search strategies for conducting knowledge- building and theory- generating qualitative systematic reviews. *Journal of Advanced Nursing*, 69(1), 194-204.
- Giddens, A. (1984). *The Constitution of Society: Outline of the Theory of Structuration*. Cambridge, UK: Polity Press.
- Glaser, B. (2002). Conceptualization: On Theory and Theorizing Using Grounded Theory. *International Journal of Qualitative Methods*, 1(2).
- Glaser, B., & Strauss, A. (1967). *The discovery grounded theory: strategies for qualitative inquiry*. Aldin, Chicago.
- Glaser, B., & Strauss, A. (1971). *Status passage*. London: Routledge & Kegan Paul.
- Green, J., & Thorogood, N. (2009). *Qualitative methods for health research* (Vol. 2nd ed.). Los Angeles, CA: Sage.
- Green, J., & Thorogood, N. (2014). *Qualitative methods for health research* (Vol. 3rd ed.). Los Angeles, CA: Sage.
- Greenhalgh, T., Robert, G., Macfarlane, F., Bate, P., Kyriakidou, O., & Peacock, R. (2005). Storylines of research in diffusion of innovation: A meta-narrative approach to systematic review. *Social Science & Medicine*, 61(2), 417-430.
- Hannes, K., & Macaitis, K. (2012). A move to more systematic and transparent approaches in qualitative evidence synthesis: update on a review of published papers. *Qualitative Research*, 12(4), 402-442.
- Harden, A., Garcia, J., Oliver, S., Rees, R., Shepherd, J., Brunton, G., & Oakley, A. (2004). Applying systematic review methods to studies of people's views: an

- example from public health research. *Journal of Epidemiology and Community Health*, 58(9), 794-800.
- Higgins, J. P., & Green, S. (2008). *Cochrane handbook for systematic reviews of interventions* (Vol. 5): Wiley Online Library.
- Houston, S. (2001). Beyond social constructionism: Critical realism and social work. *British Journal of Social Work*, 31(6), 845-861.
- Jamal, F., Fletcher, A., Harden, A., Wells, H., Thomas, J., & Bonell, C. (2013). The school environment and student health: a systematic review and meta-ethnography of qualitative research. *BMC public health*, 13(1), 798.
- Jensen, L. A., & Allen, M. N. (1996). Meta-synthesis of qualitative findings. *Qualitative health research*, 6(4), 553-560.
- Kearney, M. H. (2001). Levels and applications of qualitative research evidence. *Research in nursing & health*, 24(2), 145-153.
- Koerting, J., Smith, E., Knowles, M., Latter, S., Elsey, H., McCann, D., . . . Sonuga-Barke, E. (2013). Barriers to, and facilitators of, parenting programmes for childhood behaviour problems: a qualitative synthesis of studies of parents' and professionals' perceptions. *European child & adolescent psychiatry*, 22(11), 653-670.
- Lee, R. P., Hart, R. I., Watson, R. M., & Rapley, T. (2014). Qualitative synthesis in practice: some pragmatics of meta-ethnography. *Qualitative Research*, 1468794114524221.
- Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Ioannidis, J. P., . . . Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *Annals of internal medicine*, 151(4), W-65-W-94.
- Lofland, J. (1970). Interactionist imagery and analytic interruptus. *Human nature and collective behavior: Papers in honor of Herbert Blumer*, 35-45.
- Major, C. H., & Savin-Baden, M. (2010). *An introduction to qualitative research synthesis: Managing the information explosion in social science research*: Routledge.
- Malpass, A., Shaw, A., Sharp, D., Walter, F., Feder, G., Ridd, M., & Kessler, D. (2009). "Medication career" or "Moral career"? The two sides of managing

- antidepressants: A meta-ethnography of patients' experience of antidepressants. *Social Science & Medicine*, 68(1), 154-168.
- Malterud, K. (2001). Qualitative research: standards, challenges, and guidelines. *The lancet*, 358(9280), 483-488.
- Melendez-Torres, G. J., Grant, S., & Bonell, C. (2015). A systematic review and critical appraisal of qualitative metasynthetic practice in public health to develop a taxonomy of operations of reciprocal translation. *Research Synthesis Methods*.
- Miles, M., & Huberman, A. (1984). Qualitative data analysis: A sourcebook of new methods *Qualitative data analysis: a sourcebook of new methods*: Sage publications.
- Miles, M., & Huberman, A. (1994). *Qualitative data analysis: An expanded sourcebook*: Sage.
- Mills, A. J., Durepos, G., & Wiebe, E. (2010). *Encyclopedia of case study research* (Vol. 2): Sage Publications.
- Noblit, G. W., & Hare, R. D. (1988). *Meta-ethnography: Synthesizing qualitative studies* (Vol. 11): Sage.
- O'Mara-Eves, A., Thomas, J., McNaught, J., Miwa, M., & Ananiadou, S. (2015). Using text mining for study identification in systematic reviews: A systematic review of current approaches. *Systematic Reviews*, 4(5).
- Oliver, S., Harden, A., Rees, R., Shepherd, J., Brunton, G., Garcia, J., & Oakley, A. (2005). An emerging framework for including different types of evidence in systematic reviews for public policy. *Evaluation*, 11(4), 428-446.
- Paterson, B. L., Thorne, S., Canam, C., & Jillings, C. (2001). *Meta-study of qualitative health research: a practical guide to meta-analysis and meta-synthesis*. Thousand Oaks, CA: Sage.
- Pope, C., Ziebland, S., & Mays, N. (2000). Qualitative research in health care: Analysing qualitative data. *BMJ: British Medical Journal*, 320(7227), 114.
- Riegler, A. (2005). Editorial: The constructivist challenge. *Constructivist Foundations*, 1(1), 1-8.
- Riese, H., Carlsen, B., & Glenton, C. (2014). Qualitative Research Synthesis: How the Whole Can Be Greater than the Sum of Its Parts. *Anthropology in Action*, 21(2), 23-30.

- Riese, H., Samara, A., & Lillejord, S. (2012). Peer relations in peer learning. *International Journal of Qualitative Studies in Education*, 25(5), 601-624.
- Rolfe, G. (2006). Validity, trustworthiness and rigour: quality and the idea of qualitative research. *Journal of Advanced Nursing*, 53(3), 304-310.
- Saldana, J. (2009). An introduction to codes and coding. *The coding manual for qualitative researchers*, 1-31.
- Sandelowski, M. (1993). Rigor or rigor mortis: the problem of rigor in qualitative research revisited. *Advances in Nursing Science*, 16(2), 1-8.
- Sandelowski, M. (2012). Metasynthesis of qualitative research. In H. Cooper (Ed.), *APA handbook of research methods in psychology: Research designs: Quantitative, qualitative, neuropsychological, and biological* (Vol. 2, pp. 19-36). Washington, DC: American Psychological Association.
- Sandelowski, M., & Barroso, J. (2007). *Handbook for synthesizing qualitative research*: Springer Publishing Company.
- Sandelowski, M., Docherty, S., & Emden, C. (1997). Focus on qualitative methods
Qualitative metasynthesis: issues and techniques. *Research in Nursing and Health*, 20, 365-372.
- Schutz, A. (1962). *Collected papers* (Vol. 1). The Hague: Martinus Nijhoff.
- Shaw, R. L., Booth, A., Sutton, A. J., Miller, T., Smith, J. A., Young, B., . . . Dixon-Woods, M. (2004). Finding qualitative research: an evaluation of search strategies. *BMC Medical Research Methodology*, 4(1), 5.
- Stansfield, C., Brunton, G., & Rees, R. (2014). Search wide, dig deep: literature searching for qualitative research. An analysis of the publication formats and information sources used for four systematic reviews in public health. *Research Synthesis Methods*, 5(2), 142-151.
- Strauss, A. L. (1970). Discovering new theory from previous theory. In T. Shibutani (Ed.), *Human nature and collective behavior: Papers in honor of Herbert Blumer* (pp. 46-53). New Brunswick, NJ: Transition Books.
- Suddaby, R. (2006). From the editors: What grounded theory is not. *Academy of management journal*, 49(4), 633-642.
- Taylor, B., Wylie, E., Dempster, M., & Donnelly, M. (2007). A case study evaluating seven databases. *Research on Social Work Practice*.

- Thomas, J., & Harden, A. (2008). Methods for the thematic synthesis of qualitative research in systematic reviews. *BMC Medical Research Methodology*, 8(1), 45.
- Thorne, S., Jensen, L., Kearney, M. H., Noblit, G., & Sandelowski, M. (2004). Qualitative metasynthesis: reflections on methodological orientation and ideological agenda. *Qualitative Health Research*, 14(10), 1342-1365.
- Tondeur, J., Van Braak, J., Sang, G., Voogt, J., Fisser, P., & Ottenbreit-Leftwich, A. (2012). Preparing pre-service teachers to integrate technology in education: A synthesis of qualitative evidence. *Computers & Education*, 59(1), 134-144.
- Tong, A., Flemming, K., McInnes, E., Oliver, S., & Craig, J. (2012). Enhancing transparency in reporting the synthesis of qualitative research: ENTREQ. *BMC Medical Research Methodology*, 12(1), 181.
- Tong, A., Sainsbury, P., & Craig, J. (2007). Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. *International Journal for Quality in Health Care*, 19(6), 349-357.
- Walsh, D., & Downe, S. (2005). Meta- synthesis method for qualitative research: a literature review. *Journal of Advanced Nursing*, 50(2), 204-211.
- Zimmer, L. (2006). Qualitative meta- synthesis: a question of dialoguing with texts. *Journal of Advanced Nursing*, 53(3), 311-318.

Appendix D: Risk of bias tables for quantitative strand of Study One

Study ID	Limerick, Ireland		
Bias		Authors' judgment	Support for judgment
Recruitment	Individuals recruited after clusters assigned to intervention/control?	Low risk	Yes
Selection	Random sequence generation	Low risk	'An independent statistician used a computer generated random number sequence to randomly and blindly allocate teachers on a 1:1 basis to an intervention or waiting-list control group.' (Hickey et al., 2015, p. 5).
	Allocation concealment	Low risk	'Concealment of the allocation sequence was ensured by randomising participants after baseline assessments were completed. Participant allocation was conveyed to teachers by a member of the research team who, at post-intervention assessment, could not be blind to treatment allocation.' (Hickey et al., 2015, p. 5).
Performance	Blinding of participants and personnel	Unclear risk	'Participant allocation was conveyed to teachers by a member of the research team who, at post-intervention assessment, could not be blind to treatment allocation. All other researchers were blind to group allocation to minimise potential bias. Teachers were asked not to indicate to researchers whether they had received the training.' (Hickey et al., 2015, p. 5).
Detection	Blinding of outcome assessment	Unclear risk	'Participant allocation was conveyed to teachers by a member of the research team who, at post-intervention assessment, could not be blind to treatment allocation. All other researchers were blind to group allocation to minimise potential

			bias. Teachers were asked not to indicate to researchers whether they had received the training.' (Hickey et al., 2015, p. 5). Teachers provided reports of child behaviour, which were not possible to be collected blind to allocation.
Attrition	Incomplete outcome data – cluster level	Low risk	'In total, data were collected for 234 index children at baseline. However, 17 children (7%) were lost to follow up (see Fig 1). This yielded a total of 217 index children (107 control group children and 110 intervention group children) (102 boys and 115 girls)² with differing levels of behavioural problems, who were included in the final analysis (71 high scoring children, 74 low scoring children and 72 children who scored at, above, or below the mean depending on the distribution of scores)^{3,4}. The analyses in this report were conducted on teachers (n=22) and index children (n=217).' (McGilloway et al., 2010, p. 9).
	Incomplete outcome data – individual level	Low risk	
Statistical analysis	Incorrect analysis given clustering of data	Low risk	ICCs reported in Table 4 (Hickey et al., 2015, p. 13). 'Robust multiple linear regression was used to examine post-intervention outcomes for index children and the entire classroom, controlling for treatment group, classroom and baseline score. This approach controls for the potentially confounding effects of clustering by adjusting the standard errors; this accounts for the fact that pupils within classrooms or clusters were likely to share certain unobserved characteristics and, therefore, could not be considered independent.' (Hickey et al., 2015, p. 10).
	Intention-to-treat (ITT)	Unclear risk	'In total, data were collected for 234 index children at baseline. However, 17 children (7%) were lost

	analysis		to follow up (see Fig 1). This yielded a total of 217 index children (107 control group children and 110 intervention group children) (102 boys and 115 girls) ² with differing levels of behavioural problems, who were included in the final analysis (71 high scoring children, 74 low scoring children and 72 children who scored at, above, or below the mean depending on the distribution of scores) ^{3,4} . The analyses in this report were conducted on teachers (n=22) and index children (n=217).’ (McGilloway et al., 2010, p. 9).
Reporting	Reporting of baseline comparability of clusters	Low risk	‘Teachers in the intervention group were observed to use statistically significantly more negative behaviours than their control group counterparts at baseline. There were no other significant differences between teachers and neither were there any significant differences between the intervention and control groups on measures of child behaviour (using independent-samples t-tests and χ^2).’ (Hickey et al., 2015, p. 10-11)
	Selective reporting of outcomes	Low risk	All measures mentioned in report covered in results section.

Study ID	Kingston, Jamaica (2)		
Bias		Authors’ judgment	Support for judgment
Recruitment	Individuals recruited after clusters assigned to intervention/control?	Low risk	Yes

Selection	Random sequence generation	Low risk	'Schools were randomised to intervention or control using a computer-generated simple-randomisation sequence by an independent statistician who was masked to school identity.' (Baker-Henningham et al., 2012, p. 102).
	Allocation concealment	Low risk	'Schools were randomised to intervention or control using a computer-generated simple-randomisation sequence by an independent statistician who was masked to school identity.' (Baker-Henningham et al., 2012, p. 102). 'Children were recruited in the summer term prior to randomisation except for the 3-year-old children entering school in September who were screened and recruited after randomisation by a researcher who was masked to group allocation. Parents of these children were unaware of the school's intervention status.' (Baker-Henningham et al., 2012, p. 102-103).
Performance	Blinding of participants and personnel	Unclear risk	'Teachers were aware of allocation status at baseline and post-intervention. To maintain masking of researchers: teachers were asked at initial contact and prior to each assessment not to reveal intervention status;' (Baker-Henningham et al., 2012, p. 104).
Detection	Blinding of outcome assessment	Unclear risk	'Measurements were conducted by researchers masked to study design, hypothesis and group allocation, who conducted equal numbers of measurements per group [...] Teachers were aware of allocation status at baseline and post-intervention. To maintain masking of researchers: teachers were asked at initial contact and prior to each assessment not to reveal intervention status; observers and interviewers were not informed of the study design and were employed during the measurement phases of the study only; and intervention and control schools were provided

			with the same materials.' (Baker-Henningham et al., 2012, p. 104).
Attrition	Incomplete outcome data – cluster level	Low risk	'All 24 pre-schools were followed to the end of the trial (Fig. 1). Fifteen children were lost to follow-up (eight intervention and seven control) and these children did not differ from those retained.' (Baker-Henningham et al., 2012, p. 104).
	Incomplete outcome data – individual level	Low risk	
Statistical analysis	Incorrect analysis given clustering of data	Low risk	'The effect of intervention on continuous variables was examined using multilevel multiple regression models, which are the appropriate form of analysis for a clustered trial where outcomes are observed at level 1 (children), who are taught in classes (level 2) and the intervention is at level 3 (pre-school).' (Baker-Henningham et al., 2012, p. 104).
	Intention-to-treat (ITT) analysis	Low risk	'Analyses were conducted on an intention-to-treat basis, using the baseline scores for missing data at post-intervention.' (Baker-Henningham et al., 2012, p. 104).
Reporting	Reporting of baseline comparability of clusters	Low risk	'There were no significant differences between the groups on child, family, classroom or school characteristics at baseline (Table 1).' (Baker-Henningham et al., 2012, p. 104)
	Selective reporting of outcomes	Unclear risk	All child measures mentioned in report covered in results section, but unclear whether teacher outcomes were measured too.

Study ID	Rhode Island, USA		
Bias		Authors' judgment	Support for judgment
Recruitment	Individuals recruited after clusters	Unclear risk	Not addressed in report.

	assigned to intervention/co ntrol?		
Selection	Random sequence generation	Unclear risk	'7 teaching teams and 1 education manager were randomly assigned to receive IYS-TP.' (Shepard et al., 2008).
	Allocation concealment	Unclear risk	'7 teaching teams and 1 education manager were randomly assigned to receive IYS-TP.' (Shepard et al., 2008).
Performance	Blinding of participants and personnel	Unclear risk	Participating teaching staff were unable to be blind to allocation. 'Trained observers blind to classroom condition rated classrooms in the Fall prior to IYS-TP (Baseline), the Spring following IYS-TP (Post-Test) and the following Fall (Follow-up).' (Shepard et al., 2008).
	Blinding of outcome assessment	Unclear risk	Participating teaching staff were unable to be blind to allocation. 'Trained observers blind to classroom condition rated classrooms in the Fall prior to IYS-TP (Baseline), the Spring following IYS-TP (Post-Test) and the following Fall (Follow-up).' (Shepard et al., 2008).
Attrition	Incomplete outcome data – cluster level	Low risk	Not addressed in report.
	Incomplete outcome data – individual level	Unclear risk	
Statistical analysis	Incorrect analysis given clustering of data	Unclear risk	Not addressed in report.
	Intention-to-	Unclear	Not addressed in report.

	treat (ITT) analysis	risk	
Reporting	Reporting of baseline comparability of clusters	Low risk	‘IYS-TP classrooms had slightly lower classroom environment scores than comparison classrooms at baseline [...] At baseline, there were no significant differences between IYS-TP and comparison teachers on use of positive or inappropriate strategies, or use of positive approaches with parents. [...] At baseline, IYS-TP teachers perceived more closeness [...] and more conflict [...] with students than comparison teachers’ (Shepard et al., 2008)
	Selective reporting of outcomes	Unclear risk	Limited information available from poster format.

Study ID	North Carolina, USA		
Bias		Authors’ judgment	Support for judgment
Recruitment	Individuals recruited after clusters assigned to intervention/co ntrol?	Low risk	Yes
Selection	Random sequence generation	Unclear risk	'This randomized, controlled efficacy study is designed evaluate [sic] IYT for K-2 teachers and students in a 4-level (students, teachers, grade levels, schools) randomized block design with treatment at level 2 (teachers). Randomization to IYT or wait-list control occurs with blocking within schools such that at least one participating grade level is assigned to the intervention condition.' (Murray et al., 2012, p. 3).

	Allocation concealment	Unclear risk	Not addressed in reports.
Performance	Blinding of participants and personnel	Unclear risk	Participating teaching staff were unable to be blind to allocation. 'Following teacher consent and randomization, trained research assistants blind to randomization status observed each teacher's classroom for approximately 2 hours of instructional time, during which they rated four consecutive 30-minute cycles of classroom activities and summarized the overall classroom environment on a 7-point Likert scale [...]' (Murray et al., 2012, p. 3).
Detection	Blinding of outcome assessment	Unclear risk	Participating teaching staff were unable to be blind to allocation. 'Following teacher consent and randomization, trained research assistants blind to randomization status observed each teacher's classroom for approximately 2 hours of instructional time, during which they rated four consecutive 30-minute cycles of classroom activities and summarized the overall classroom environment on a 7-point Likert scale [...]' (Murray et al., 2012, p. 3). Teacher outcomes are all based on blind observations (low risk), while student outcomes are based on teacher reports (unclear risk).
Attrition	Incomplete outcome data – cluster level	Low risk	'Participants were 44 K-2 teachers [...]. Of these 44 teachers, data were collected at two time points from 38.' (Murray et al., 2012, p. 2). Attrition not addressed in 2014 report.
	Incomplete outcome data – individual level	Unclear risk	
Statistical analysis	Incorrect analysis given	Unclear risk	Not addressed in report.

	clustering of data		
	Intention-to-treat (ITT) analysis	Unclear risk	Not addressed in report.
Reporting	Reporting of baseline comparability of clusters	Low risk	See section on baseline equivalence in Murray et al., 2014 report.
	Selective reporting of outcomes	Unclear risk	Preliminary findings from a study so not all outcomes reported.

Study ID	North Wales		
Bias		Authors' judgment	Support for judgment
Recruitment	Individuals recruited after clusters assigned to intervention/control?	Low risk	Yes
Selection	Random sequence generation	Low risk	'This study adopted a single-blind stratified RCT design. An independent researcher paired classrooms according to school size, classroom size, and locality (town/rural) after preintervention measures had been collected. Schools were subsequently randomly allocated to the intervention or control condition using a random number generator: the paired school was automatically allocated to the control group in each case.' (Hutchings et al., 2013, p. 577).
	Allocation	Low risk	'This study adopted a single-blind stratified RCT

	concealment		design. An independent researcher paired classrooms according to school size, classroom size, and locality (town/rural) after preintervention measures had been collected. Schools were subsequently randomly allocated to the intervention or control condition using a random number generator: the paired school was automatically allocated to the control group in each case.' (Hutchings et al., 2013, p. 577).
Performance	Blinding of participants and personnel	Unclear risk	'Observations at preintervention were undertaken prior to randomization, and observers were blind to both children's TSDQ scores and teachers' group allocation (intervention or control) at postintervention [...] This study adopted a single-blind stratified RCT design.' (Hutchings et al., 2013, p. 577).
Detection	Blinding of outcome assessment	Low risk	'Observations at preintervention were undertaken prior to randomization, and observers were blind to both children's TSDQ scores and teachers' group allocation (intervention or control) at postintervention.' (Hutchings et al., 2013, p. 577).
Attrition	Incomplete outcome data – cluster level	Low risk	'All teachers were retained for the duration of the study.' (Hutchings et al., 2013, p. 574).
	Incomplete outcome data – individual level	Low risk	'Analyses were conducted using full maximum estimation likelihood to account for missing data under an assumption that data were missing at random. Only one child in the control group was lost to follow-up (1%). Complete data were available for the TSDQ TOTAL and PRE SCORES covariates: however, there were missing data for the GENDER (4%) and AGE (4%) covariates.' (Hutchings et al., 2013, p. 578).
Statistical analysis	Incorrect analysis given clustering of	Low risk	'To examine the effect of intervention on teacher and child behavior, hierarchical linear modeling (HLM) analyses were conducted using SPSS 17.0

	data		to account for the clustered nature of the data.' (Hutchings et al., 2013, p. 577). ICCs reported in results tables.
	Intention-to-treat (ITT) analysis	Low risk	Analyses were run on all participants, and there was no attrition.
Reporting	Reporting of baseline comparability of clusters	Low risk	‘There were no significant differences in child, teacher, or school demographics between control and intervention groups preintervention (Table 1). There were no significant differences in the number of children in the low and high TSDQ groups between the intervention and control conditions preintervention, $\chi^2(1) = 0.03, p = .868$.’ (Hutchings et al., 2013, p. 580)
	Selective reporting of outcomes	Low risk	All measures mentioned in report covered in results section.

Appendix E: Quality appraisal tables for qualitative strand of Study One

Study ID	Devon, England	
Bias	Authors' judgment	Support for judgment
Rigor of sampling	High quality	All participating headteachers interviewed. Researchers sampled more than 80% of participating teachers (31 out of 37 who completed the course). It is unclear why the missing six teachers did not participate in the interviews. ‘Between two and four teachers from 16 schools attended the course, with places allocated on a “first come, first served” basis. [...] The selection of teachers from within each school was decided by the head. [...] places were open to teachers from any primary school year group, and participants were teaching classes that ranged from Reception (age 5) to Year 5 (age 10) (see Table 1). As is typical for the teaching profession, most teachers recruited were women, but the sample had a broad range of teaching experience, and included teachers in their newly qualified year as well as special educational needs co-ordinators’ (Marlow et al., 2015, p. 233). ‘In total, 40 teachers were invited to attend TCM training one day per month for six months, and 37 teachers completed the course. Of the three teachers who did not complete the course, one left the teaching profession, one chose to leave the course after two sessions feeling it was too basic and one attended three sessions but was then unable to attend further sessions due to staffing issues.’ (Marlow et al., 2015, p. 233). ‘In total, 31 teachers attended focus groups, which were led by a researcher-facilitator, with an observer also present.’ (Marlow et al., 2015, p. 235). ‘Headteachers were invited to participate in a semi-structured interview over the telephone after their staff member had completed the course. All the headteachers from participating

		schools were interviewed.’ (Marlow et al., 2015, p. 235).
Rigor of data collected	Medium quality	Nothing mentioned about piloting topic guides etc., although this was part of a feasibility study, so it was piloting the process for the main study. ‘The [focus] groups ran for a maximum of one-and –half [sic] hours with clear ground-rules (including confidentiality) and structure. The topic guide (available from authors on request) included questions about course content, style and mode of delivery, theory, practical skills, materials, group learning, arrangements for attending and feasibility, dissemination within their school and other programmes used within schools. It also included questions about their experience of the research process including recruitment and questionnaires, and suggestions for improvements to inform the main trail. Each group concluded with a period of summary and reflection and participants had the opportunity to discuss the focus group with a researcher via e-mail should they wish, though none chose to do so.’ (Marlow et al., 2015, p. 235). ‘Headteachers were invited to participate in a semi-structured interview over the telephone after their staff member had completed the course. All the headteachers from participating schools were interviews. Interviews lasted a maximum of 45 minutes and followed a topic guide (available from authors on request) that included background information about the school, feedback from teachers about the TCM course and feasibility of attendance and how TCM compares with any other similar programmes within their school.’ (Marlow et al., 2015, p. 235).
Rigor of analysis	Medium quality	‘Focus group discussions and headteacher interviews were audio-recorded, transcribed and anonymised. Data were stored using NVivo 9 software. The thematic frameworks used to code the teacher focus groups and headteacher interviews are illustrated in the following lists; the coding frameworks were shaped by the topic guides but also developed iteratively to include emergent themes.’ (Marlow et

		al., 2015, p. 235). ‘Analysis drew on understanding of social learning [...] and was approached from a subtle realist perspective, identifying experiences as the lived “reality” of participants [...] and exploring the ways in which participants account for their experiences within the context of the study and their own schools [...] The Framework approach [...] was used to manage data and aid the systematic description and summary of key themes, patterns and links in the data.’ (Marlow et al., 2015, p. 238). ‘In the analysis, “keyness” of themes did not necessarily relate to incidence of occurrence but to whether a theme captured information relevant to the research questions [...] in this case relating to a range of experiences and views about behaviour and context relevant to the TCM intervention.’ (Marlow et al., 2015, p. 238). Diversity of perspective was not directly addressed in the report. No description of seeking out alternative explanations or descriptions of contradictory cases.
Findings grounding in data	High quality	Enough data are presented to show how the authors arrived at their findings; data presented fit the interpretation and illustrate the findings. See throughout the findings section. Quotations are not numbered or otherwise identified so that the reader can see that they do not just come from one or two people.
Breadth and depth	Medium quality	Decent breadth on a range of issues but not much depth in the findings. Beyond presenting the perspectives of the participants and using a quotation to illustrate the point, there is little exploration of contrasting perspectives or providing insight into the perspectives. Richness and complexity do not come through due to lack of depth. Not really any theoretical or conceptual development.
Overall reliability and trustworthiness	Medium quality	Based on sampling most of the available teachers and all headteachers and grounding their findings in the data, this study is reliable and trustworthy. However, because the report

		does not address possibly contradictory cases or seeking out dissenting voices and does not explore the perspectives of the researchers analysing the data, it is not rated as 'high'.
Overall usefulness of findings for this review	High quality	The study aims and findings are a very good match for the purpose of the synthesis, even with some lack of conceptual depth.

Study ID	Greater Dublin Area, Ireland	
Bias	Authors' judgment	Support for judgment
Rigor of sampling	Medium quality	All teachers participating in a pre-test/post-test study on IY TCM were asked to complete a qualitative survey. '15 teachers, selected on the basis of first response to advertisement of an Incredible Years Teacher Training Course, attended a five month programme consisting of two consecutive afternoons (3.30 pm to 6.30 pm) per month at an Education Centre in the Greater Dublin Area from February to June 2007. Schools represented were from both suburban and rural settings.' (Davenport & Tansey, 2009, p. 4). No characteristics of the sample were described in the report, limiting readers' understanding of the study context and findings that were presented.
Rigor of data collected	Low quality	Nothing mentioned about piloting topic guides etc. Nothing mentioned about how comprehensive or sensitive the survey was, or how complete/vivid/rich people's descriptions were. The qualitative survey was administered with the other quantitative measures as part of the overall study, and as such it is expected that participants felt able and willing to contribute; nothing is mentioned suggesting otherwise and

		there were no face-to-face meetings to highlight potential power relations etc.
Rigor of analysis	Low quality	Analysis of qualitative data was minimally described: ‘Qualitative results from the evaluation questionnaire were grouped and listed according to frequency of occurrence under the respective headings.’ (Davenport & Tansey, 2009, p. 6). Diversity of perspective was not directly addressed in the report. The analysis did not seek to rule out alternative explanations for findings or to interpret the findings; it just reported the findings.
Findings grounding in data	Low quality	No quotations or other data presented to show how the authors arrived at their findings.
Breadth and depth	Low quality	While a range of topics are mentioned, they are not discussed sufficiently to be ‘covered’ and there is no depth explored. The perspectives of the participants are not explored beyond a frequency count described typically as ‘all’ or ‘most’. This account is not rich or complex. There is no theoretical or conceptual development.
Overall reliability and trustworthiness	Low quality	This study samples all of the participating teachers on the IY TCM course using a qualitative survey, rather than interview or focus group methods. However, the very little detail available in the methods, analysis, and findings sections severely limit its reliability and trustworthiness. No qualitative data is presented to support the findings, and all qualitative findings are very briefly presented in a few short sentences.
Overall usefulness of findings for this review	Low quality	The succinct and undescriptive vote-count reporting of the qualitative data in this report results in low usefulness of this study for the review. Although the types of topics mentioned could have been relevant to the review, they are not interpreted or presented with any sort of depth to encourage further consideration within the collection of qualitative data on this

		IY TCM programme.
--	--	-------------------

Study ID	Limerick, Ireland	
Bias	Authors' judgment	Support for judgment
Rigor of sampling	High quality	‘Although all RCT participants were provided with information about the qualitative interviews, maximum variation purposive sampling was used to obtain a sample of 12 teachers differing in age, service years, school types, and school locations. [...] Initially, it was intended that all those who had been interviewed at baseline would be re-interviewed at follow-up time points. However, three teachers who took part in baseline interviews were unavailable at follow-up. Due to a non-permanent teaching post, one teacher (T4) was no longer working in the school at the 6-month follow-up (control group) Another (T10) was ineligible for participation at that time due to a move to a resource teaching post whilst prior personal commitments resulted in the unavailability of a third teacher (T11). At 12-month follow-up, similar challenges in re-interviewing participants emerged; one (T9) was on sick leave and another (T21) (due to non-permanency of post) was no longer working in her original school. [...] To address this shortfall in numbers, three teachers who had not previously participated in the qualitative research process, were interviewed at the 12-month follow-up; these participants were selected solely due to their availability and willingness to be involved.’ (Hyland, 2014, p. 110). ‘[...] the sample in the qualitative study was selected to be as diverse as possible and comprised teachers, principals, TCM group facilitators, and the Archways development manager.’ (Hyland, 2014, p. 109)
Rigor of data collected	High quality	‘The researcher developed a series of interview schedules [...] for teachers, principals, programme facilitators, and the

		development manager, respectively. Interview content was informed by a review of the literature and by the research questions’ (Hyland, 2014, p. 112). ‘Interviews with teachers and principals were held insofar as possible, in quiet rooms on school premises, typically a vacant classroom or the principal’s office. These interviews ranged in duration from 30 to 70 minutes, and all were recorded on microcassette. [...] Interviews with the development manager and one TCM facilitator were held at Archways, Clondalkin. The other facilitator interview was held in a secondary school in West Dublin. All interviews with Archways personnel lasted approximately 60 minutes.’ (Hyland, 2014, p. 116). ‘The aim during the interviews was to balance adherence to pre-set questions with attention to emergent issues. The interviewer was aware of potential bias in terms of question selection and in leading the general direction of the interviews. Therefore, a number of opportunities were given to the interviewees to voice their opinion on other issues they perceived as relevant. Follow-up and probing questions were also used to obtain additional information. Trust and rapport were easily established because the researcher had been involved in the early recruitment of teachers to the trial, and had met with them (and the principals) on several occasions prior to interviews. Similarly, the positive working relationship between the researcher and programme delivery staff enabled open and frank interview discussions.’ (Hyland, 2014, p. 116).
Rigor of analysis	High quality	‘Qualitative interviews were audio-recorded and transcribed verbatim to Microsoft Word. Although qualitative analysis software (MAXQDA) was trialled as a possible means of facilitating data storage/analysis, data were subsequently stored using Microsoft Excel and all analysis was conducted using hard copies of transcripts.’ (Hyland, 2014, p. 119).

		‘A number of analytical methods could have been used in the qualitative study, but it was decided that Framework Analysis provided the ‘best fit’ with the aims and objectives of the study.’ (Hyland, 2014, p. 116). ‘According to Charmaz (2006), no analysis is neutral and there is always the potential for contamination. [...] For the above reasons, the researcher engaged in a process of reflection throughout the study’ (Hyland, 2014, p. 128). ‘Despite the predetermined, deductive nature of framework identification, it was vital that predetermined aims of the study were implemented alongside an inductive approach towards the data. Thus, this framework which incorporated both a priori and emergent issues, developed in a recursive [...] rather than linear process, throughout the analytic process.’ (Hyland, 2014, p. 122). The analysis did not specifically address contradictory cases or seek to rule out alternative explanations for findings, although differences in experiences were mentioned.
Findings grounding in data	High quality	Enough data are presented to show how the authors arrived at their findings, the data presented fit the interpretation and illuminate the findings. Quotations are numbered to anonymously identify participants so that readers can see that the quotations do not just come from one or two people.
Breadth and depth	High quality	A range of issues are covered; the perspectives of participants are fully explored in terms of breadth and depth; richness and complexity have been portrayed and there has been some theoretical and conceptual development.
Overall reliability and trustworthiness	High quality	Across the three reports related to this study (and most particularly with the Hyland, 2014 report) substantial detail regarding the approach to research, methods, analysis, and findings are communicated to rate this study high in terms of reliability and trustworthiness of its findings.

Overall usefulness of findings for this review	High quality	Due to the relevant topics covered in this study and the considerable depth of the reported findings, this study rates high in terms of its usefulness of its findings for this review.
---	--------------	---

Study ID	Kingston, Jamaica (1)	
Bias	Authors' judgment	Support for judgment
Rigor of sampling	High quality	Sampled all participating teachers except one teacher who left study early. 'The schools consisted of 15 classrooms staffed by 16 teachers. One teacher left the school before the end of the study. The remaining 15 teachers (13 women, two men) were interviewed for this study.' (Baker-Henningham & Walker, 2009, p. 633). Sampled all participating teachers who completed study. But, no explanation for why one teacher left school/study early. Also addressed lack of resources to extend interviews to parents but encourage this as an area for future research. (Baker-Henningham & Walker, 2009, p. 640). Minimal information shared about attempts to obtain a diverse sample, but as they describe the teachers' backgrounds as: 'The remaining 15 teachers (13 women, two men) were interviewed for this study. Twelve teachers had completed a basic vocational course in early childhood care, education and development but only one was a trained teacher.' (Baker-Henningham & Walker, 2009, p.633).
Rigor of data collected	Medium quality	Nothing mentioned about piloting topic guides etc. 'At the end of the intervention period, in-depth individual interviews with each teacher were conducted by a research assistant who had not been involved with the study. Interviews were conducted in a quiet location at school and

		they lasted between 45 min and 1 h. A semi-structured interview guide was used which covered the following topics: 1) teachers' overall opinion of the training programme and their perspectives of any benefits of the training to themselves and to the children; 2) how teachers used the new strategies learned in the classroom, most and least useful components of the training and the difficulties encountered implementing the strategies; 3) teachers' views on how the intervention led to changes in their own and the children's behaviour; 4) recommendations for future implementation of the training programme.' (Baker-Henningham & Walker, 2009, p.633-634). 'Although a researcher who had not been involved in the intervention conducted the interviews, it is possible that teachers enhanced the benefits and minimized the difficulties associated with the intervention during the in-depth interview.' (Baker-Henningham & Walker, 2009, p. 640).
Rigor of analysis	Medium quality	See analysis section p. 634 for description of Framework Approach etc. '[...] it was stressed that information was required to improve the intervention and teachers' accounts were rich with specific examples of how they used the strategies and specific details of the reported behavioural changes.' (Baker-Henningham & Walker, 2009, p. 640). It is unclear what preconceptions researchers had, but data was driving the analysis as per analysis description and grounding findings in direct quotations to illustrate points. No description of seeking out alternative explanations or descriptions of contradictory cases. But, there is a section on difficulties, see p. 636.
Findings grounding in	High quality	Enough data are presented to show how the authors arrived at their findings; data presented fit the interpretation and

data		illuminate the findings – see throughout findings section. No numbering or other identification system used for quotations to demonstrate that they do not come from just one or two people.
Breadth and depth	Medium quality	Good breadth on a range of issues but not much depth in the findings. Beyond presenting the perspectives of the participants and using a quotation to illustrate the point, there is little exploration of contrasting perspectives or providing insight into the perspectives. Richness and complexity do not come through due to lack of depth. Not really any theoretical or conceptual development.
Overall reliability and trustworthiness	High quality	Based on sampling all available teachers and grounding their findings in the data, this study is reliable and trustworthy.
Overall usefulness of findings for this review	High quality	The study aims and findings are a very good match for the purpose of the synthesis, even though the conceptual depth is somewhat lacking.

Study ID	North Wales (1)	
Bias	Authors' judgment	Support for judgment
Rigor of sampling	High quality	‘The 23 teachers were nominated by their employers (two Authorities in North West Wales). They came for a variety of reasons, including teacher interest and employer identified need. [...] Interviews were undertaken with 21 teachers by the third author.’ (Hutchings et al., 2007, p. 19). Any attempts to obtain a diverse sample are not addressed in the report, and it is unclear why the two teachers out of 23 participating in the training did not take part in the interviews. Characteristics of the entire RCT sample are presented, and the

		interview participants are a sub-sample of that.
Rigor of data collected	Unclear quality	Nothing mentioned about piloting topic guides etc. Insufficient information about interviewing process to determine whether data collection was comprehensive, flexible, and sensitive enough to provide a complete/vivid/rich description of people's perspectives. Nothing mentioned about steps taken to ensure that all participants were able and willing to contribute.
Rigor of analysis	Unclear quality	'[...] qualitative analysis was conducted on the transcripts of interviews using thematic content analysis. These were coded into seven categories by two independent coders. To measure inter-rater reliability, Kappa's coefficient was calculated and agreement was found to be high ($\kappa = 0.82$).' Hutchings et al., 2007, p. 19). Diversity in perspective was not directly addressed in the report. The extent to which the analysis was guided by preconception or by the data was not directly addressed in the report. The analysis presented different views but did not seek to rule out alternative explanations for findings or search for negative cases.
Findings grounding in data	Medium quality	A fair amount of data are presented to show how the authors arrived at their findings and the data presented fit the interpretation. The data illustrate the findings fairly well, although there is no numbering or other identification system used for quotations to demonstrate that they do not come from just one or two people.
Breadth and depth	Medium quality	A wide range of issues are mentioned, but there is very little depth across these topics. There is no richness or complexity portrayed in the findings and not really any theoretical or conceptual development in the report.
Overall	Medium	In terms of the reliability and trustworthiness of the findings

reliability and trustworthiness	quality	from this study, there is just too little detail provided on the approach to research, methods, and analysis to rate this study more than 'medium'.
Overall usefulness of findings for this review	Medium quality	Overall, the topics covered are relevant and useful for this review, but due to the lack of depth provided in the reported findings, this study is rated 'medium' for its usefulness of its findings for this review.

Appendix F: SENCo topic guide for Study Two

Supporting Teachers And children in Schools Main Phase – Impact of course post-intervention

Special Educational Needs Coordinator (SENCo) - Topic Guide for Telephone Interview to take place June/July 2014 (C02); June-July 2015 (C03)

[More detailed interview guides will be developed to assist researchers and maintain consistency during data collection, to include: introductions, equipment checks, parameters and ground rules, brief explanation of interview process, roles and timing]

Pre-interview questions

- How long have you been a teacher? Are you an in-class SENCo?
- For how many years have you worked in your current school and how long have you served as the SENCo?
- What is the percentage of children with SEN in your school, and what is the most common type of SEN?
- On average, how severe do you perceive the level of SEN is in your school?

General SEN and behaviour management

1. In your experience, what classroom management strategies are successful for children with SEN who have behavioural problems?
Further prompt (if necessary): What unique classroom management strategies would you use when working with children with SEN who have behavioural problems?
2. How do you, as SENCo, assist teachers of children with SEN who have special behavioural needs?
Further prompt (if necessary): Describe the last time a teacher came to you to discuss a child with SEN who was exhibiting problematic behaviours—what was the problem and how did you both address it?

Incredible Years Teacher Classroom Management (IY TCM) Program (STARS Study)

3. During this academic year, one of the teachers in your school attended a classroom management course called The Incredible Years (STARS Study). What, if anything, have you heard about it?
Further prompt (if necessary): Has [teacher's name] discussed attending a classroom management training course? If so, what did [teacher's name] say about it?

If they have heard about the IY TCM course ask Question #4; if not, skip to Question #5.

4. Based on what you have heard from [teacher's name] about the strategies taught on the course, what do you perceive are the benefits for children with SEN who have special behavioural needs and where might the training fall short?
Further prompt (if necessary): How do you think the children with SEN who also exhibit problematic behaviours responded to the strategies taught on the Incredible Years course? How might teachers need to adapt these management strategies when working with children with SEN who have additional behavioural needs?

SEN services in the school

5. What internal/external support services are currently available in your school for you and other teachers when supporting children with SEN who have special behavioural needs?
Further prompt (if necessary): What additional resources/specialist services can your school offer to support children with SEN who have behavioural problems?

Wrap-up

6. Is there anything that I haven't asked that you would like to add—any comments that you feel are important to this conversation?

Supporting Teachers And childRen in Schools

Information Sheet for Telephone Interviews with Special Educational Needs Coordinators (SENcos)

Background to STARS Study

If children enjoy their early years at school and settle into good learning habits, it gives them a really good start that can stay with them for years. Our research study is about how children and teachers get on together in school. In many classrooms, there are children whose behaviour is difficult for the teacher to manage, and this can be disruptive for the rest of the class and a great source of stress for some teachers.

You may be aware that your school is currently taking part in the STARS study. The study is exploring whether a Teacher Classroom Management (TCM) course can help teachers deal with children who are disruptive in class. This course is thought to help teachers to support these children and thus reduce disruption to the rest of the class, leading to a general improvement in all their pupils' socio-emotional skills.

One teacher from your school is taking part in the study and has recently attended the TCM course. We would like to find out whether you, as SENCo, have noticed any impact in the school as a result of the teacher attending the TCM course.

What would I be asked to do?

We are inviting you to take part in a short telephone interview to discuss your views about SEN support in general and any impact of the TCM course in your school.

Interviews are useful for collecting relevant and detailed information on people's views on a particular topic or experience. Conducting interviews by telephone is a convenient way of collecting this information without too much imposition on people's time. Telephone interviews with SENcos will take place shortly after the participating teachers have

finished their TCM course. A doctoral student researcher will conduct these interviews. This researcher, Elizabeth Nye, is a special education teacher with classroom experience teaching primary school pupils with special educational needs (SEN).

What information is being collected and what use will be made of it?

The purpose of these interviews with the SENCOs is to learn about the unique aspects of working with pupils with SEN who also have special behavioural needs, and whether the TCM course fits in with existing support services offered through the school for pupils with SEN. This information will add to the existing data collected as part of the STARS study.

What will happen during the telephone interview?

The researcher will ring you at a pre-arranged convenient time to carry out the interview. Before starting the interview, the researcher will verbally request consent for participation and to audio record the discussion. Audio recording the interview allows the researcher to transcribe the discussion so that subsequent analysis of participants' responses is accurate. These transcripts will be used to analyse the data to describe commonalities and differences in SENCOs' experiences and views.

The researcher will follow a guide with a list of topics and questions during this interview, ensuring that all SENCOs are asked about the same topics. The researcher is interested in *your* views and experiences in relation to students with both SEN and special behavioural needs, support services in the school, and any impact of the TCM course. The researcher will ask questions and provide time for your responses, and you will have an opportunity to make suggestions about additional issues. You can refuse to answer any questions that you do not wish to. The interview will last around 30 minutes.

Do I have to take part?

Taking part in this telephone interview is completely voluntary, and there will be no disadvantage to you of any kind if you do not wish to participate.

What happens now?

One of the research team will email you to see whether you would like to take part in the interview. If you would like to, we will ask you to respond by email to say that you are willing to take part. We will then put you in touch with Elizabeth who will arrange a convenient time with you to carry out the interview.

Gratitude

In appreciation of the time given to this research, if you decide to take part you will receive a £10 gift voucher following the interview.

Ethical approval

This study has been reviewed and approved by the Peninsula College of Medicine and Dentistry Research Ethics Committee, now under the auspices of the University of Exeter Medical School Research Ethics Committee.

Complaints

If you have any complaints about the way in which this study has been carried out please contact the Chair of the University of Exeter Medical School Research Ethics Committee:

Peta Foxall, PhD

Chair, UEMS Research Ethics Committee

Email: P.J.D.Foxall@exeter.ac.uk

Any questions?

If you have any questions or concerns about any aspect of this project, please speak to the doctoral student researcher Elizabeth Nye (elizabeth.nye@spi.ox.ac.uk; 07585 773267) who will do her best to answer your query.

The STARS study is led by Professor Tamsin Ford (T.J.Ford@exeter.ac.uk; 01392 722973), and the Trial Manager is Rachel Hayes (R.A.Hayes@exeter.ac.uk; 01392 722978), both of whom can answer any questions or respond to any concerns you might have. You may also contact Elizabeth's doctoral supervisor, Professor Frances Gardner (frances.gardner@spi.ox.ac.uk; 01865 270325) at the University of Oxford.

Thank you for considering this study.

Appendix H: Standard operating procedure for consent for Study Two

Standard Operating Procedure: Obtaining consent over the phone

We have approval from UEMS Ethics to obtain participants' consent over the telephone when this conversation is being audio recorded (e.g. telephone interviews with teachers, SENCOs etc.).

Once contact has been made with the participant (either by email or by telephone) send them a copy of the appropriate information leaflet by email:

For teachers - *Telephone Interview Information Leaflet v1*

For SENCOs/TAs – *Information Sheet for Telephone Interviews with Special Educational Needs Coordinators (SENCOs) and Teaching Assistants (TAs) v2*

Arrange a convenient time for the telephone interview.

At the start of the interview

To gain a participant's consent over the phone, first introduce yourself and what the interview will be about and that it will be audio recorded. Then follow the following script:

Would you be happy for me to start the tape recording equipment now?

So, before we start we need to gain your consent to take part in this interview. To do this, please could you confirm that you are happy with and understand the following:

- You understand that your participation in this telephone interview is entirely voluntary;
- You understand that you do not have to answer any questions you don't want to and that you are free to stop the interview at anytime;
- You understand that the interview will be recorded and that this audio recording will remain in secure storage;
- You understand that the results of the study may be published but any data we collect from you today will remain anonymous at all times.
- You understand that there will be an opportunity to revisit your consent at the end of the interview and you can change your mind if you would like to.

Is that all ok? Do you have any questions?

So, one final thing to confirm, are you happy to take part in the interview and for me to continue now with the first question?

This verbal consent will be recorded in the interview transcript.

Appendix I: Ethics approval for Study Two

DEPARTMENT OF SOCIAL POLICY AND INTERVENTION

Barnett House, 32 Wellington Square,
Oxford, OX1 2ER, United Kingdom
www.spi.ox.ac.uk



Elizabeth Nye
Department of Social Policy and Intervention
Oxford

Ref: 2013-14_56

02 June 2014

Dear Elizabeth,

Perspectives of special educational needs coordinators on working with students who have both special educational needs and special behavioural needs

Your application for research ethics approval in connection with your thesis has been considered by the Departmental Research Ethics Committee (DREC) in accordance with the procedures laid down by the University for Ethical Approval.

I am pleased to inform you that, on the basis of the information provided, the proposed research has been judged as meeting appropriate ethical standards and DREC approval has been granted.

If any revisions to your research methodology are made subsequent to this approval, these must be detailed in writing and submitted to DREC immediately.

Yours sincerely,

A handwritten signature in black ink that reads 'Lucie Cluver'.

Dr Lucie Cluver
Chair of DREC